



Redes sociales y comunicación publicitaria.

La metodología SETAL para la obtención
de *insights* publicitarios.

Xavier Duran Muñoz

Director: Dr. Josep A. Rom Rodríguez

Tesis doctoral | Barcelona, noviembre 2015



TESIS DOCTORAL

Título : Redes Sociales y comunicación publicitaria.
La metodología SETAL para la obtención de *insights*
publicitarios.

Realizada por : Xavier Duran Muñoz

en el Centro : Facultad de Comunicación Blanquerna

y en el Departamento: Comunicación

Dirigida por: Dr. Josep A. Rom Rodríguez

A l'Esther, la Raquel, en Dani i en Marc

Agraïments

Molt content d'arribar a la part corresponent als agraïments, perquè això vol dir que he arribat a la fi d'aquest treball de tesi doctoral i perquè sempre és agradable agrair a la molta gent que m'ha ajudat al llarg d'aquest procés.

Si, ha estat un camí molt llarg, poder massa, ja que els estudis de doctorat els vaig començar el 2006. Vaig acabar la tesina el 2009 i vaig fer la seva presentació davant del tribunal de la DEA el 2010. Del 2010 al 2012 vaig començar un treball de recerca d'una tesi doctoral que vaig abandonar la primera setmana de gener del 2013, és a dir, porto gairebé tres anys d'ençà que vaig començar a treballar en aquesta nova tesi doctoral.

Haig de dir que, durant tot aquest període, la conciliació de la meua vida professional, al davant de la meua agència Objectivos Web, les classes com a professor del departament de Publicitat de la Facultat de Comunicació i Relacions Internacionals Blanquerna, i la vida familiar, no ha estat fàcil. A més, el darrer any, he estat nomenat Director General de la Fundació Sitges - Festival Internacional de Cinema de Catalunya. Un gran honor i un gran repte que he hagut de compatibilitzar, per trobar temps de qualitat necessari per poder concentrar-me en el desenvolupament d'aquesta tesi. Ha estat un gran esforç que en aquests moments torno a viure amb una estranya i ambigua sensació d'alegria, alleugeriment i angoixa.

Precisament, perquè el recorregut ha estat llarg i feixuc, hi ha hagut moltes persones, professionals i companys que m'han ajudat en aquest projecte i que cal que rebin el meu més sincer agraïment.

Començaré pel principi i aniré avançant en el temps. Així doncs el meu primer agraïment és pel Dr. Miquel Altarriba qui es va proposar convertir a un professional de la comunicació en un professor i em va indicar el camí a seguir en el perfeccionament i professionalització acadèmica. El Dr. Altarriba, va ser el meu primer director de la primera fase de la tesi doctoral fins que es va jubilar.

Vull recordar al Pare Andreu Marqués, monjo de Montserrat que ens va deixar el juny del 2012. Recordo els passejos pel Camí dels degotalls i els seus consells. Va ser el meu pare espiritual i qui em va inculcar l'esperit del sacrifici per a assolir les fites sense agafar dreceres. Ell va ser un gran professor, un gran mestre. Un savi que va insistir en la importància de fer la tesi doctoral i vull que el seu record quedi present en aquest treball, que ha estat el treball més important que mai he fet fins ara, dins l'àmbit acadèmic.

Com deia abans, hi va haver un moment en què vaig deixar enrere tres anys de treball del primer projecte de tesi doctoral. Cada any, el primer cap de setmana de gener, tinc una cita anual amb el professor Joost van Nispen, president de l'Institut de Comerç Electrònic i Màrqueting Directe (ICEMD). Durant el dinar anual, gràcies a la seva experiència i en un moment de reflexió en veu alta, vaig veure clar que aquell objecte d'estudi mancava d'actualitat i interès. Gràcies Joost per ajudar-me a veure-hi clar.

Així doncs, l'actual projecte de tesi doctoral va agafar un nou rumb, que no hagués estat possible sense la immensa col·laboració, imprescindible, per poder tirar endavant aquest projecte del Dr. Enric Monte, professor de la Universitat Politècnica de Catalunya. Els requeriments tecnològics necessaris per fer el resum i l'anàlisi dels textos no hauria estat possible sense el desenvolupament de l'eina SETAL (Sentiment Text Analysis). Amb ell he treballat intensament en aquest procés que ara sembla arribar a la fi. Moltes gràcies Enric.

Aquest nou projecte de tesi, ha estat dirigit pel Dr. Josep A. Rom, qui va substituir al Dr. Altarriba i qui amb la seva rapidesa i agudesia, identificava ràpidament els punts de millora i els possibles paranys que sempre van apareixent al llarg d'aquest camí. Vull agrair la col·laboració del professor Enric Mur pel seu assessorament en el camp de la investigació de mercats i validació estadística de les mostres. Vull agrair la col·laboració dels meus exalumnes Sergi Garcia i Gil Padrol així com d'en Sergio Espinosa.

També vull agrair la col·laboració de tots els professionals, directors dels departaments de comunicació digital de les empreses entrevistades pel treball de recerca qualitativa, que em

van obrir les portes de les seves empreses i departaments sense limitacions per donar-me tot tipus d'informació per aquesta tesi. Gràcies a l'Andreu Sans de eLaCaixa, al Ferran Pallàs del Banc Sabadell-Atlantico, a la María José Martínez de Catalunya Caixa, a Saúl Rodríguez del BBVA, a Rafel Feliu de Danone i a la Amets Uriarte de Nestlé.

Finalment, vull agrair a la meva família i a la meva dona l'Esther Ciberio pel seu suport. L'Esther ha estat la persona que m'ha donat l'equilibri i la tranquil·litat necessària per poder treballar i concentrar-me en la tesi. Als meus fills, perquè ells són l'estímul, la meva força i la meva vida.

A tots i totes, moltes gràcies.

Sitges, setembre de 2015.

Resumen

El objeto principal de estudio de esta tesis doctoral se basa en la investigación y desarrollo de una nueva metodología para la elaboración de *insights* publicitarios, realizada a partir del análisis netnográfico de las conversaciones que los usuarios realizan de forma libre y espontánea en la red social Twitter (en nuestro ámbito de investigación y por razones de capacidad y conveniencia, lo hemos limitado a la red social Twitter).

A partir del análisis de la literatura académica hemos documentado las técnicas de investigación mayormente utilizadas por los investigadores para la elaboración de *insights* publicitarios. Dichas técnicas se basan en entrevistas y reuniones de grupos reducidos de entrevistados (habitualmente con menos de diez), dirigidos por un entrevistador. Hemos analizado sus ventajas e inconvenientes valorando los estudios y trabajos de referencia en el ámbito académico y concluido que los sesgos producidos por las propias dinámicas de las reuniones de grupo y de la interpretación de los resultados por parte de los entrevistadores suponen un inconveniente importante.

Con el objetivo de superar estas limitaciones, tanto de representatividad como de sesgo, hemos investigado las posibilidades que el entorno de Internet y las redes sociales pueden aportar para la obtención de una fuente de información más amplia, representativa y sin sesgo. Por un lado, hemos sustituido las reuniones de grupo por las conversaciones que realizan miles de personas en Twitter. Por otro lado, eliminamos la figura del entrevistador por una investigación de análisis de contenido mediante la técnica netnográfica, donde el proceso de investigación se basa en la observación del consumidor sin intervención directa del entrevistador, eliminando prácticamente el sesgo que ocasiona su presencia.

Para poder alcanzar este objetivo principal de investigación se ha requerido el desarrollo de una herramienta específica de resumen y análisis de textos, tanto desde el análisis gramatical como de los *sentimientos* expresados por los usuarios en sus comunicaciones.

Hemos utilizado la herramienta “SETAL”, acrónimo de *Sentiment Text Analysis*. SETAL es una nueva herramienta de resumen y análisis gramatical y semántico de textos en fase beta, desarrollada por el Dr. Enric Monte, profesor de la Universidad Politécnica de Cataluña, probada adecuadamente para esta tesis doctoral.

Por una lado, SETAL ordena todos los contenidos por palabras, oraciones, textos e incluso usuarios, según su orden de importancia, elaborado por la coocurrencia de palabras clave, las propiedades de la Teoría de Grafos y de su distribución estadística (similar al algoritmo utilizado por el motor de búsqueda de Google para determinar el PageRank de los sitios web, a partir de una cadena de Markov). Hemos dedicado una parte de esta tesis a la explicación del funcionamiento de SETAL.

Por otro lado, hemos incorporado a la investigación gramatical y semántica la variable *sentimiento* expresado por los usuarios en sus comentarios en Twitter. Nuestra intención ha sido querer superar el análisis de la polaridad positiva, negativa o neutra y pasar a la precisión que permite puntuar cada oración en función de la expresión de los sentimientos (dentro de un sentido amplio detallado en el punto 4 de esta tesis).

Para ello hemos utilizado como punto de partida el diccionario LIWC (*Linguistic Inquiry and Word Count*), desarrollado por el Dr. James Pennebaker et al., y adaptado al castellano y catalán para esta tesis. Hemos partido de este diccionario para puntuar las palabras que forman las oraciones y que reconocidos grupos de investigación en psicología han determinado que están asociadas con una determinada intensidad de sentimientos/emociones. Añadimos al análisis de sentimientos la utilización de los adverbios y los tiempos gramaticales, así como los bigramas (combinaciones de dos palabras cuya carga emocional se potencia a partir de su confluencia).

Mediante la prueba realizada con la herramienta SETAL hemos dado respuesta a la gestión de 1.200.000 tuits y nos ha resultado imprescindible para poder gestionar y analizar de forma automática un volumen de información tan elevado.

Al poner en práctica la herramienta siguiendo un proceso propio amparado por las referencias académicas, proponemos una nueva metodología para la elaboración de *insights*. Su principal aportación puede resumirse en tres apartados. En primer lugar, ampliamos considerablemente la muestra del estudio superando con creces las limitaciones de las entrevistas presenciales y pasamos de un máximo recomendado de diez entrevistados para ese tipo de entrevistas a poder estudiar varios cientos de personas con nuestra metodología. En segundo lugar, logramos eliminar el sesgo dada la no intervención de la figura del entrevistador en el proceso de obtención de la información. En tercer lugar, proponemos una serie de datos obtenidos de forma científica que nos permiten reducir o eliminar la dependencia de la intuición, el descubrimiento súbito o el "¡ajá" asociado al hallazgo de un *insight* (explicado detenidamente en las partes teóricas de esta tesis).

En resumen, esta tesis doctoral busca dar un paso más en la dirección que nos permite elaborar contenidos publicitarios con la solidez que aporta partir de una base científica y objetiva. Para lograrlo, esta visión científica ha sido vinculada al potencial que ofrece el análisis de los contenidos que los usuarios crean y desarrollan en Internet, concretamente en las redes sociales, y que nos aportan una novedosa forma de obtener información fundamental acerca de todo aquello que les importa y deciden compartir. Para una empresa, estar lo más cerca posible del día a día del consumidor a través de las redes sociales supone una oportunidad de vital importancia para seguir ofreciéndole todos los productos, servicios y comunicaciones que necesita, con la mejor comunicación posible y pudiendo demostrarle que se le escucha al ser su bienestar la mayor prioridad y cometido de toda labor empresarial.

Abstract

The main object of study of this doctoral thesis is the research and development of a new methodology for the improvement of advertising insights, through the netnographic analysis of the conversations that users of the social network Twitter perform freely and spontaneously (we have limited our field of study to Twitter for capacity and operational reasons).

From the analysis of academic literature we have documented the investigation techniques researchers mostly use for the development of advertising insights. Such techniques are based on interviews and meetings of small groups of individuals (usually less than ten), led by an interviewer. We have analyzed their pros and cons evaluating the academic studies of reference and concluded that the biases produced by the dynamics of the group meetings and the interpretation of the results by the interviewers pose a major drawback.

In order to overcome these limitations, regarding biases and representativeness, we have investigated the possibilities that the field of Internet and Social Media can contribute with towards the obtention of a broader, more representative and unbiased source of information. On one hand, we have replaced the group meetings by the conversations that thousands of people have on Twitter. On the other we have removed the interviewer, and have substituted its figure with a content analysis research of netnographic technique where the research process is based on the observation of consumers without the direct intervention of the interviewers, therefore virtually eliminating biases caused by their presence.

The achievement of this main objective of research has required the development of a specific tool for the summary and analysis of texts, both grammatical and emotional, that users express in their communications.

We have used "SETAL", which stands for *Sentiment Text Analysis*. SETAL is a new tool for grammatical and semantic summary analysis of texts in phase beta, developed by Dr. Enric Monte, professor of the Polytechnic University of Catalonia, and adequately tested for this thesis.

On one hand, SETAL organized all content by words, sentences, texts and even users, according to their importance, elaborated by the concurrence of keywords, the properties of the Graph Theory and its statistical distribution (similar to the algorithm used by the search engine Google to determine the PageRank of websites, from a chain of Markov). We have devoted a part of this thesis in explaining the operation of SETAL.

On the other, we have incorporated to the grammatical and semantic research the emotion variable expressed by users in their comments on Twitter. Our aim has been to take further the analysis of positive, negative or neutral polarity and reach an accuracy that scores each sentence on the basis of the expression of feelings (broadly detailed in paragraph 4 of this thesis).

So we've used as a starting point the LIWC (*Linguistic Inquiry and Word Count*) dictionary developed by Dr. James Pennebaker et al., and adapted it to Castilian and Catalan for this thesis. We have started from this dictionary to score the words and sentences that recognized psychology research groups have identified as associated with certain intensity of feelings / emotions. We add to Sentiment Analysis the use of adverbs and grammatical tenses as well as bigrams (two-word combinations whose emotional power is increased by its confluence).

Through the SETAL test we have managed 1,200,000 tweets. The tool has proved essential to manage and automatically analyze such an important volume of information.

In implementing the tool following a proper process under academic references, we propose a new methodology for the development of insights. Its main contribution can be summarized in three sections. First, we significantly expand the study sample, far exceeding the limitations of face interviews and its recommended maximum of 10 subjects, to study several hundred people through our methodology. Second, we eliminate bias thanks to the non presence of the interviewer in the process of obtaining information. Thirdly, we propose a series of scientific data that reduces or eliminates the reliance on intuition, the sudden discovery or "aha" associated with the finding of an insight (that we carefully explain in the theoretical part of this thesis).

In summary, this thesis seeks to go a step further in the direction that allows us to develop advertising content with the confidence provided by a scientific and objective basis. To achieve this, this scientific vision has been linked to the potential of the analysis of the content that users create and develop online, particularly on social networks, and we provide a novel way to obtain basic information about all that matters to them and that they decide to share. For a company, to be as close as possible to the daily life of consumers through social networks is of vital importance. To keep on offering all products, services and releases, with the best possible communication with them as a way to prove that they are listened, as their wellbeing is the highest priority and commitment of any business.

ÍNDICE

PRESENTACIÓN DEL TRABAJO DE INVESTIGACIÓN	25
CONTRIBUCIÓN PRINCIPAL	31
OBJETIVOS E HIPÓTESIS	35
METODOLOGÍA	41
TERMINOLOGÍA	51
ESTRUCTURA Y PARTES DEL TRABAJO DE INVESTIGACIÓN	67
PRIMERA PARTE	69
1. MARCO TEÓRICO Y CONCEPTUAL	73
1.1. LA COMUNICACIÓN EN LOS MEDIOS SOCIALES	73
1.1.1. Comunicación de masas y medios de comunicación	75
1.1.2. Los Medios Sociales (<i>Social Media</i>)	76
1.1.3. Los Medios sociales son Medios de comunicación de masas	86
2. LA INVESTIGACIÓN EN LA COMUNICACIÓN PUBLICITARIA	91
2.1. EL INSIGHT EN LA COMUNICACIÓN PUBLICITARIA	91
2.1.1. El <i>insight</i> creativo	99
2.1.2. El <i>consumer insight</i>	100
2.1.3. Aclaraciones acerca de los <i>insights</i>	103
2.1.3.1. Ejemplos de <i>insights</i>	105
2.2. MÉTODOS DE INVESTIGACIÓN APLICADOS A LA COMUNICACIÓN	109
2.2.1. Métodos tradicionales de investigación	110
2.2.1.1. Métodos cuantitativos	110
2.2.1.1.1. La encuesta	111
2.2.1.1.2. El cuestionario	112
2.2.1.1.3. Análisis de contenido	112
2.2.1.1.4. Pre-test, post-test y estudios de seguimiento continuo	114
2.2.1.2. Métodos cualitativos	116
2.2.1.2.1. La entrevista	117
2.2.1.2.2. La entrevista en profundidad	117
2.2.1.2.3. La entrevista telefónica	118
2.2.1.2.4. La entrevista por correo	118
2.2.1.2.5. La entrevista por Internet	118
2.2.1.2.6. Focus Group o “Grupo Focal”	120
2.2.1.2.7. El método Delphi	121
2.2.1.3. Cuadro de resumen de los métodos tradicionales de investigación	122
2.2.1.4. Conclusión acerca de los métodos de investigación tradicionales	123
2.2.2. Técnicas proyectivas de investigación	126
2.2.3. Técnicas específicas de generación de <i>insights</i>	129

3. TEORÍA DE GRAFOS	137
3.1. ANTECEDENTES	138
3.2. DEFINICIÓN Y COMPONENTES DE UN GRAFO	162
3.2.1. Definición	162
3.2.2. Componentes de un grafo	163
3.3. APLICACIONES DE LA TEORÍA DE GRAFOS	172
3.3.1. Aplicaciones por sectores de actividad	174
3.3.2. Tipos de redes	174
3.3.2.1. Redes tecnológicas	175
3.3.2.2. Redes sociales	177
3.3.2.2.1. Análisis de Redes Sociales (ARS)	179
3.3.2.3. Redes de información	183
3.3.2.4. Redes biológicas	184
4. RESUMEN AUTOMÁTICO DE TEXTOS (RAT)	191
4.1. ANTECEDENTES DEL RESUMEN AUTOMÁTICO DE TEXTOS	194
4.2. DEFINICIÓN	197
4.3. TAXONOMÍA Y MÉTODOS	201
4.3.1. Taxonomía de los métodos del RAT	202
4.3.1.1. RAT según la forma	203
4.3.1.1.1. Resumen extractivo	203
4.3.1.1.2. Resumen abstractivo	203
4.3.1.2. RAT según el propósito	204
4.3.1.2.1. Resumen indicativo	204
4.3.1.2.2. Resumen informativo	204
4.3.1.2.3. Resumen crítico	204
4.3.1.3. RAT según la audiencia	204
4.3.1.3.1. Genérico	204
4.3.1.3.2. Query-based	205
4.3.1.3.3. RAT “guiados” hacia un objetivo	205
4.3.1.4. RAT según la cantidad de documentos a resumir	205
4.3.1.5. RAT según el idioma	206
4.3.1.5.1. Plurilingüe	206
4.3.1.6. RAT según las fuentes	206
4.3.1.6.1. Multimedia	206
4.3.1.7. RAT según sus métodos	206
4.3.1.7.1. Métodos estadísticos	206
4.3.1.7.2. Métodos basados en la conectividad del texto (Text Connectivity)	207
4.3.1.7.3. Métodos de Grafos Iterativos	208
4.3.1.7.4. Método de la Lógica Difusa (Fuzzy Logic)	208
4.3.2. Técnicas de evaluación de resúmenes	209
4.3.2.1. Evaluación Intrínseca	210
4.3.2.2. Evaluación Extrínseca	216
4.3.2.2.1. Sistemas de evaluación extrínseca	217
4.3.3. Detección de la subjetividad y los sentimientos	218
4.3.3.1. Clasificación del sentimiento en Internet	219
4.3.3.2. Clasificación del sentimiento de productos	220
4.3.3.3. Clasificación del sentimiento en frases (textos)	220
4.3.3.4. Clasificación de Facetas del Sentimiento	221
4.3.3.5. SOCal (<i>Semantic Orientation Calculator</i>)	222

4.3.4. Uso de grafos en los resúmenes automáticos de textos	222
4.4. APLICACIONES DEL RESUMEN AUTOMÁTICO DE TEXTOS	224
4.4.1. Resumen de correos electrónicos	228
4.4.2. Resumen de Blogs y Chats	230
4.4.3. Resumen de mensajes en Twitter	232
5. SENTIMENT ANALYSIS	239
5.1. ANTECEDENTES	241
5.2. DEFINICIÓN	244
5.2.1. Emociones, sentimientos y opiniones	246
5.2.1.1. Emoción	247
5.2.1.2. Sentimientos	252
5.2.1.3. Opiniones	252
5.3. TIPOLOGÍA DE MÉTODOS DEL SENTIMENT ANALYSIS	254
5.3.1. Tipología de métodos basados en el <i>Machine Learning</i> (ML)	255
5.3.1.1. Naïve Bayes	255
5.3.1.2. Maximum Entropy	255
5.3.1.3. Support Vector Machines (SVM)	256
5.3.2. Tipología de métodos basados en el léxico (<i>lexicon-based</i>)	256
5.3.2.1. Cálculo de Orientación Semántica (Semantic Orientation Calculator) (SOCal)	256
5.3.2.2. WordNet	257
5.3.2.3. MPQA Subjectivity Lexicon	258
5.3.2.4. SenticNet	258
5.3.2.5. Linguistic Inquiry and Word Count (LIWC)	259
5.3.3. Los diferentes niveles de clasificación en el Sentiment Analysis	260
5.3.3.1. Nivel del texto (documento) completo	260
5.3.3.2. Nivel de las frases	261
5.3.3.3. Nivel del aspecto de las palabras	261
5.4. CAMPOS Y POSIBILIDADES DE APLICACIÓN DEL SENTIMENT ANALYSIS (SA)	262
5.4.1. Aplicaciones del SA en Twitter	263
SEGUNDA PARTE	267
6. SETAL (SENTIMENT TEXT ANALYSIS)	271
6.1. DESCRIPCIÓN DE SETAL	272
6.2. METODOLOGÍA SETAL PARA LA OBTENCIÓN DE INSIGHTS	275
6.3. CONSIDERACIONES DEL FUNCIONAMIENTO DE SETAL	276
6.3.1. Fase de obtención de la información	279
6.3.1.1. Ámbito de interés	279
6.3.1.2. Identificación de los influencers	279
6.3.1.3. Obtención de la información	280
6.3.1.4. Filtros: eliminación del ruido	280
6.3.2. Fase de análisis de la información	281
6.3.2.1. Análisis gramatical de los tuits	281
6.3.2.2. Análisis de los Sentimientos (Sentiment Analysis)	281
6.3.3. Fase de resumen de la información	282
6.3.3.1. Jerarquización y resumen de toda la información	282
6.3.4. Fase de elaboración del <i>insight</i>	284

7. INVESTIGACIÓN EMPÍRICA**291****7.1. CONSIDERACIONES****294**

7.1.1. Respecto a la metodología

294

7.1.2. Respecto al uso de la netnografía

294

7.1.3. Respecto a la privacidad de los contenidos estudiados

295

7.1.4. Respecto al uso de la folsconomía

296

7.1.5. Respecto al análisis gramatical de los contenidos

297

7.1.5.1. Palabras funcionales (*Function Words*)

297

7.1.5.2. Palabras de contenido (*Content words*)

299

7.1.6. Respecto al uso del diccionario *Linguistic Inquiry and Word Count (LIWC)*

299

7.2. IDENTIFICACIÓN DE LA MUESTRA**301**

7.2.1. Selección de la red social

302

7.2.1.1. Análisis de Facebook

302

7.2.1.2. YouTube

302

7.2.1.3. Google Plus

302

7.2.1.4. Twitter

302

7.2.2. Selección del sector de actividad

304

7.2.3. Variables técnicas de análisis

305

7.2.4. Resumen de los textos de los tuits

307

7.2.5. Medición de los sentimientos de los tuits

307

7.3. INVESTIGACIÓN EMPÍRICA**310**

7.3.1. Proceso de análisis de los contenidos

310

7.3.1.1. Criterios de clasificación de las variables de análisis

310

7.3.1.2. Análisis de las variables de estudio

312

7.3.1.2.1. Clasificación temporal de los sentimientos

312

7.3.1.2.2. Oraciones más importantes

316

7.3.1.2.3. Palabras más importantes

319

7.3.1.2.4. Segmentación por *topics*

322

7.3.1.2.5. Distribución porcentual de los tuits por *topics*

326

7.3.1.2.6. Ejemplos de tuits según sus *topics*

327

7.3.1.2.7. Análisis de los usuarios por *topics*

330

7.3.1.2.8. Usuarios más importantes

334

7.3.1.2.9. Sentimientos asociados a los usuarios

336

7.3.1.3. Análisis gramatical de los contenidos

338

7.3.1.4. Ejemplo de aplicación de la metodología “SETAL” para la elaboración de *insights*

340

8. INVESTIGACIÓN CUALITATIVA**349****8.1. METODOLOGÍA DE LA INVESTIGACIÓN CUALITATIVA****351**

8.1.1. La entrevista en profundidad

354

8.1.1.1. Modelo de la entrevista en profundidad

354

8.1.1.2. Resultado de las entrevistas en profundidad

357

9. CONCLUSIONES	369
10. BIBLIOGRAFÍA	381
ANEXOS	401
DICCIONARIO LIWC	401
DICCIONARIO DEL SECTOR BANCARIO	415
TRANSCRIPCIÓN DE LAS ENTREVISTAS	425
TABLA DE LOS DIEZ PRIMEROS USUARIOS POR ORDEN DE IMPORTANCIA	453
ÍNDICE DE TABLAS	455
ÍNDICE DE FIGURAS	457

PRESENTACIÓN DEL TRABAJO DE INVESTIGACIÓN

PRESENTACIÓN DEL TRABAJO DE INVESTIGACIÓN

Si revisamos y estudiamos las cifras del número de usuarios activos en las redes sociales alrededor del mundo en este 2015, la cuestión ya no es si su presencia permanecerá en un futuro o si es solo una tendencia temporal. Tras los últimos estudios realizados acerca del uso de dos de las mayores redes sociales (Facebook y Twitter), medios como el Time corroboraban que una parte creciente de sus usuarios los utiliza como fuente de información, especialmente Twitter. (Grodén, 2015)^[1]

En marzo de 2015, Facebook tenía 1.450 millones de usuarios activos y Twitter rondaba los 288 millones.^[2] Ante este inédito volumen de tráfico y vista la importancia que sus propios usuarios le otorgan a la información que se comparte cotidianamente, el interés desde el ámbito empresarial por intentar obtener datos e información relevante sobre sus usuarios (gustos, opiniones, valoraciones) ha crecido también exponencialmente. Al mismo tiempo, Facebook y Twitter procuran ejercer, para rentabilizar sus plataformas, como canales de venta directos que vinculen al usuario con el proceso de compra de los productos destacados que él ha señalado como interesantes anteriormente.

Con estas sinergias, el vínculo entre el usuario como consumidor y productor de información y el usuario como consumidor de productos y servicios se ha estrechado de tal modo dentro de las mismas redes que, desde el campo de la publicidad, es imperante avanzar en el desarrollo de vías para poder acercarse a la audiencia con las comunicaciones más adecuadas a sus gustos, inquietudes y valoraciones.

Para la publicidad, por lo tanto, las oportunidades que estos nuevos medios brindan son considerables: desarrollar campañas, determinar *targets* o lanzar productos específicos han sido, entre otras posibilidades, focos de atención del marketing digital.

No obstante, en esta investigación nos centraremos en un aspecto concreto de las oportunidades que dichas redes brindan al trabajo publicitario. Utilizando Twitter como fuente de información, trataremos de obtener datos útiles, relevantes y concretos a través de

las conversaciones que los usuarios mantienen en sus comunicaciones en la red social en referencia a determinadas marcas. La voz de los usuarios, desde la expresión de sus deseos de compra hasta sus valoraciones que envían a través de tuits a las marcas que siguen (o que les desagradan), pasando por la opinión de los productos que consumen (también en el ámbito del consumo cultural y los viajes) o a las quejas concretas por la calidad del producto que han comprado, todo ello puede ser usado en beneficio de la marca que sabe escuchar: los publicistas pueden mejorar sus futuras comunicaciones publicitarias si son capaces de recopilar la información pública que circula entre usuarios y marcas y que tiene lugar en los muros de Twitter. Como hemos apuntado, a través de Twitter y haciendo uso de ciertas herramientas pensadas para ello, intentaremos mostrar que, en efecto, las redes sociales son una oportunidad para afinar y mejorar tanto los productos como los mensajes publicitarios. Confiamos en que, si todo este trabajo se realiza rigurosamente, podamos obtener *insights* publicitarios: saber qué dicen, hacen o piensan nuestros clientes (potenciales y actuales), o aquellos personajes públicos a cuya sensibilidad queremos acercarnos, es el paso clave para poder avanzarnos a la solución de las necesidades que de ahí se derivan. Sin intervención por parte de los publicistas, es decir, siendo su presencia totalmente “invisible” en dichas conversaciones, el valor del contenido estudiado obtiene todavía más entereza. Y este es el interés capital de este trabajo: convertir el ruido de la red en algo audible y manejable para realizar el trabajo publicitario con mayor eficacia, precisión y pertinencia ante los súbitos cambios de las nuevas tecnologías y sus posibilidades.

En nuestro estudio, los campos del saber que han sido explorados y a los que pertenece la información obtenida son los que enumeramos a continuación:

- Sociología
- Psicología
- Antropología
- Matemáticas
- Marketing y Comunicación

Tal y como se puede apreciar en la bibliografía utilizada, la mayoría de autores que han investigado en el entorno de comunidades sociales, redes sociales y técnicas para la

explotación de la información que se desarrolla en ellas son, principalmente, norteamericanos. Sin embargo, la presencia de autores españoles es notable y sus producciones en el ámbito investigativo son cada vez mayores. De las fuentes españolas hemos obtenido material bibliográfico, artículos académicos y tesis doctorales. Señalamos algunos de los autores y autoras más destacados:

- Elena Lloret, Óscar Ferrández, Rafael Muñoz y Manuel Palomar de la Universidad de Alicante, por su labor de investigación en la implicación textual en las tareas automáticas de resúmenes de textos.
- Virginia Francisco Gilmatín, de la Universidad Complutense de Madrid, por su tesis doctoral sobre la “Identificación automática del contenido afectivo de un texto y su papel en la presentación de la información”.
- Alexandra Balahur Dobrescu, de la Universidad de Alicante, por su tesis doctoral “*Methods and Resources for Sentiment Analysis in Multilingual Documents of Different Text Types*”.
- Cristina Balanzó, de la Facultad de Comunicación Blanquerna de la Universidad Ramón Llull, en su tesis doctoral “*Los consumer insights y las neurociencias*”.
- Cristina Martorell Castellano, de la Facultad de Comunicación Blanquerna de la Universidad Ramón Llull, en su tesis doctoral “*Segueix-nos a Facebook! La participació dels consumidors a les comunitats de marca virtuals*”.

Hemos separado el trabajo de campo en dos partes principales: la investigación cuantitativa y la investigación cualitativa. Consideramos que la parte más importante del trabajo de investigación ha sido la obtención de la información directamente de las cuentas de Twitter de usuarios activos, relacionados con unas palabras clave y que han utilizado en sus tuits.

El tiempo que hemos dedicado a extraer los datos en el trabajo de campo cuantitativo ha sido de tres meses. La información se ha obtenido directamente de 4.543 usuarios de Twitter con cuentas activas (de una segmentación inicial de 5.000 usuarios). De su procedencia se han almacenado alrededor de 100.000 tuits semanales siguiendo una serie de criterios que

detallaremos en la cuarta parte de este trabajo. Para el análisis global final se ha trabajado con un total de 1.200.000 tuits.

En cuanto a la fase de investigación cualitativa, hemos considerado necesario ponernos en contacto con representantes destacados de empresas anunciantes con una alta presencia en las redes sociales. Creemos indispensable acercarnos a su conocimiento obtenido por las exigencias de su labor acerca del uso que los usuarios hacen de las redes sociales, así como la opinión que se deriva de su experiencia profesional. Además, las empresas a las que acudimos prestan especial atención al estudio de lo que hemos convenido en denominar “Sentimientos” y en cómo los usuarios los expresan en las redes sociales. En el proceso de trabajo realizado gracias a la colaboración de las empresas consultadas, han resultado especialmente valiosas las extensas entrevistas realizadas a los directivos y a los responsables de las áreas de marketing digital y a los *community managers* que gestionan las diversas redes sociales. Las entrevistas han sido realizadas a grandes empresas de diversos sectores para obtener una visión lo más amplia posible sobre las variaciones que los profesionales del campo tienen acerca de la variable “Sentimientos”. Al apreciar las diversas variaciones en la percepción hemos podido comprender con mayor precisión el diseño y la aplicación de sus estrategias de marketing y comunicación.

Ha sido muy interesante realizar una serie de entrevistas en profundidad con los directivos y responsables de las áreas de marketing digital, *community managers* o responsables de las redes sociales de grandes empresas de diferentes sectores empresariales, para así, tener una visión lo más amplia posible de este sector y de su percepción sobre el estado actual de uso de las variables “Sentimientos” en su análisis y posterior uso para sus estrategias de marketing y comunicación.

CONTRIBUCIÓN PRINCIPAL

Contribución principal

El avance en el estudio del comportamiento de los usuarios de las redes sociales en los últimos años ha sido abismal. Para los propios usuarios, la capacidad de convertirse en emisores de contenidos y el poder compartir información sin apenas inversión económica es inédita a nivel histórico. A nivel empresarial, como hemos apuntado, la posibilidad de acercarnos a datos tan concretos acerca del comportamiento, los gustos, los hábitos y la personalidad de los usuarios nunca había podido ser tan precisa.

La conciencia por parte de las empresas que poseen las redes sociales del valor de todo ello es absoluta, de ahí que desde las propias plataformas proporcionen cada vez mayores facilidades y opciones a sus clientes para realizar campañas publicitarias. Al mismo tiempo, como veremos en este trabajo, han nacido potentes herramientas externas cuya eficacia y posibilidades expanden, complementan y mejoran en ciertos aspectos las herramientas de las propias redes sociales.

Anteriormente, en otras épocas, el método utilizado por la publicidad para obtener *insights* requería de la participación activa del publicista o los investigadores: estudios de mercado basados en encuestas, cuestionarios o dinámicas de grupo, aún siendo útiles, hacían conscientes a los participantes de que estaban siendo estudiados y ello, inevitablemente, producía un sesgo. Esa propia conciencia de los sujetos podía modificar el comportamiento “natural” de los participantes y, al mismo tiempo, no había lugar para la espontaneidad. Si bien dichos estudios admiten cierta flexibilidad, existen unos límites que no admiten mayor amplitud. Por contra, la naturalidad y autenticidad de las conversaciones en Twitter es absoluta, puesto que su temática, tono y contenido es decidido en todos los aspectos por el emisor, siguiendo los códigos comunicativos propios de la red social.

En este sentido, el avance de las redes sociales supone un nuevo paradigma para la publicidad, ya que ahora tiene los mecanismos efectivos para desarrollar sus campos de saber con métodos puramente científicos. Para este trabajo, será fundamental avanzar en este aspecto y contribuir a su demostración. Como apuntábamos, más allá de los estudios de

mercado u alternativas de otras épocas, ahora podremos dar nuevos pasos con rigurosidad científica gracias a la precisión de los programas que se han desarrollado aprovechando las nuevas informaciones y herramientas de computación.

Al realizar el trabajo de campo que presentaremos posteriormente queremos poner en evidencia que la publicidad requiere herramientas cada vez más sofisticadas para la elaboración de sus campañas de comunicación. El *Big Data* debe ser tratado sistemáticamente de modo que los datos concretos obtenidos sean válidos para ser estudiados en nuestro campo, porque si no son tratados carecen de utilidad para nuestro cometido. Las oportunidades que brinda el análisis del *Big Data* para la publicidad es, por lo tanto, otro aspecto de enorme relevancia que aspiramos a dejar patente en esta investigación. Como hemos apuntado, con los resultados obtenidos podremos mejorar el conocimiento de los consumidores y con ello nos dirigiremos a ellos de forma más adecuada.

La mayor característica de Twitter como medio de comunicación es su limitación en las comunicaciones a 140 caracteres. Asimismo, demostraremos que en un máximo de 140 caracteres puede obtenerse información relevante y un grado elevado de conocimiento de los consumidores. En dichos textos, hallaremos que los sentimientos son detectables y, además, quedará expuesto que tienen una importancia muy alta dentro del conjunto de otros elementos que podremos estudiar en los mensajes.

Es imperante remarcar que el método presentado podrá ejercer como complemento a los métodos utilizados anteriormente, y ejercerá como nueva vía disponible de forma permanente para las agencias de publicidad y con un coste relativamente bajo. Las propias empresas podrán beneficiarse también del mismo, especialmente en su modo de seguir comunicándose adecuadamente con los consumidores.

OBJETIVOS E HIPÓTESIS

Objetivos e hipótesis

Desarrollaremos el conocimiento publicitario siguiendo un método científico que pueda ser útil tanto para las empresas como para los mismos publicistas. Este objetivo será accesible usando el *Big Data* de una forma determinada y que nos permita analizar varios aspectos relevantes del consumidor en sus comunicaciones.

Al realizar la investigación del consumidor a través de sus conversaciones en Twitter, esperamos encontrar una vía tan válida como las demás para obtener *insights* publicitarios. Con ello, al mismo tiempo, demostraremos la importancia creciente que tienen los medios sociales como forma de expresión de las opiniones y sentimientos de los consumidores y de la necesidad de las empresas de conocer y utilizar esta información para la toma de decisiones, tanto a nivel comunicativo como de imagen de marca.

Google utiliza un algoritmo para jerarquizar las oraciones de las páginas web (título, entradilla, etiquetas, etc.) que luego ofrece como resultados de búsqueda en un orden que se adecua a la oración que ha buscado el usuario. El primer resultado, se considera como el más relevante, en relación al criterio de búsqueda del usuario. Sostenemos que, de forma similar, es posible organizar las comunicaciones de Twitter jerárquicamente usando herramientas de resumen automático y, por lo tanto, trabajar con ellas a posterioridad.

Dichas herramientas poseen, además, una mayor precisión: son capaces de valorar, mediante un análisis semántico, diversos factores y variables que permiten determinar, por ejemplo, la polaridad de las comunicaciones (positiva, neutra, negativa). Más allá de eso, es posible discernir, a partir de la temática del mensaje, las oraciones y el vocabulario, el estado de ánimo del usuario al realizar la comunicación o su agrado o desaprobación ante el tema del que habla.

Entendemos que la inmensa mayoría de comunicaciones acerca de productos, servicios, marcas o empresas realizadas en Twitter por usuarios con un perfil asociado a una persona física muestran efectivamente las opiniones *verdaderas* que tiene la persona acerca de ellos,

una opinión que bien puede expresar con sus compañeros de trabajo, amigos o familiares en su vida cotidiana. Twitter, en este sentido, ejerce de altavoz y nuevo canal por el que hacer llegar el mensaje a otro público, ya sean usuarios, las propias empresas o personas interesadas con una cuestión vinculada a ellos.

Con este planteamiento, herramientas y análisis de los contenidos, creemos que es posible lograr *insights* publicitarios.

Objetivo 1. Demostrar que la selección y análisis de la información que se puede obtener en las redes sociales, en concreto a través de la plataforma Twitter, es una fuente de información que puede ser utilizada para la elaboración de *insights* publicitarios.

Hipótesis 1. El algoritmo desarrollado específicamente para esta tesis doctoral, denominado SETAL acrónimo de *Sentiment Text Analysis*, que permite la obtención de información relevante del usuario y de la nueva metodología creada con el mismo nombre, se pueden elaborar *insights* publicitarios.

Las comunicaciones de los usuarios, organizadas y jerarquizadas debidamente, pueden permitir que la dirección tome ciertas decisiones en función de la opinión mostrada por la mayoría de la comunidad de Twitter. La transparencia es uno de los mayores valores que a día de hoy se exige a las empresas y, por ello, la cercanía y respuesta a las demandas, consultas, quejas o sugerencias de los usuarios han pasado a convertirse en una de las prioridades de muchas de ellas.

Como hemos apuntado en nuestros objetivos, queremos evidenciar que la dirección podrá utilizar dicha información para la toma de decisiones y, por otro lado, el departamento de comunicación y marketing podrá a su vez lanzar sus mensajes (notas de prensa, promociones, estrategia de los *community managers*) basándose también en los contenidos que ofrecen las herramientas. Los sentimientos, las emociones o las reacciones ante determinados acontecimientos de los consumidores o el público en general son ahora accesibles de primera mano, sin intermediarios, prácticamente en tiempo real y con un coste muy bajo. En la misma línea, la información podrá ser procesada para mejorar el producto o servicio, corregir errores, diferenciarlo o acercarlo al de otras empresas.

Si todo ello es posible será gracias a que las herramientas informáticas son adecuadas y pertinentes, sino las mejores para lograr dichos objetivos. Nuestra segunda hipótesis, por lo tanto, es demostrar que determinadas herramientas de computación son idóneas para gestionar, organizar y filtrar grandes cantidades de datos y cuya interpretación genera conclusiones satisfactorias.

Objetivo 2. Demostrar que, a través de Twitter y, por extensión, de las redes sociales, se puede obtener información que también sirva para los profesionales de la comunicación publicitaria y los departamentos de marketing de las empresas.

Hipótesis 2. Las herramientas de resumen automático que utilizaremos ofrecen unos resultados que son útiles, además de para los publicitarios, para otros departamentos de marketing y comunicación de las empresas.

Internet y el auge de las redes sociales aumentan su capacidad para obtener información y de establecer una comunicación directa con sus consumidores, así como su necesidad de conocer los comentarios, percepciones y manifestaciones que los consumidores hacen de sus productos y servicios.

En la actualidad hay una gran expectativa y curiosidad, por parte de los anunciantes, en conocer estas percepciones de sus consumidores y concretamente la variable “Sentimiento” expresado por sus consumidores a través de las redes sociales, pero las empresas están dentro de un proceso de integración de todo el procedimiento de análisis, validación, estudio e integración de esta variable dentro de sus planes de comunicación, del que se observa una proyección creciente.

Los anunciantes siempre han necesitado y tenido una gran curiosidad por conocer más y mejor a sus consumidores y su comportamiento en el mercado con respecto a sus productos y los de la competencia.

Objetivo 3. Investigar el estado del estudio, análisis y utilización de la variable *Sentiment* de los consumidores en las estrategias y acciones de comunicación de los anunciantes.

Hipótesis 3. La variable *Sentiment* se tiene cada vez más en cuenta para la elaboración de las estrategias de marketing y comunicación de las empresas.

METODOLOGÍA

Metodología

Para satisfacer las exigencias de la investigación, hemos utilizado en los estudios tanto el método científico hipotético-deductivo como el método analítico-interpretativo, propio de las ciencias sociales y, en concreto, utilizado las técnicas de la netnografía y la folcsonomía. El trabajo realizado con dichos métodos quedará ampliamente explicado y detallado en los apartados 4 y 5 de esta tesis.

La elección de ambos métodos nos permitirá trabajar con mayor amplitud de miras e incorporar tanto la visión de las ciencias empíricas como la herramienta más utilizada en las ciencias sociales. Con la aplicación del método hipotético-deductivo lograremos reunir datos que cumplen las condiciones del experimento científico, realizaremos hipótesis a través de su estudio y, con todo ello, deduciremos determinadas consecuencias que, en caso de ser las previstas, harán la hipótesis no falsable. Con no falsable nos referimos a que, si la seguimos poniendo a prueba y sigue cumpliéndose según el método propone, la tomamos como válida para nuestra investigación.

El método analítico interpretativo observa la realidad como algo que no puede fragmentarse sistemáticamente, y centra su estudio en lo que entendemos como “fenómenos”, que no son fácilmente aislables en un espacio-tiempo. El objeto de esta visión es que nos permitirá interpretar hechos cuya relación puede parecer más lejana o tangencial, pero que sin embargo puede resultar cabal. La subjetividad del investigador está más presente, puesto que su criterio a la hora de seleccionar y relacionar los datos posee mayores posibilidades que el estudio empírico. Por ello, uno y otro método producen conocimientos diferentes entre sí y, en nuestro caso, de valor complementario.

En cuanto al aspecto teórico que usaremos para el análisis e interpretación de las conversaciones de los usuarios con el fin de elaborar *insights*, hemos tomado tres herramientas troncales que presentamos también con un recorrido desde sus orígenes hasta la actualidad. En primer lugar, aplicaremos la Teoría de Grafos. En segunda instancia usaremos

el Resumen Automático de Textos (RAT) y por último el “*Sentiment Analysis*”. Las tres especialidades se encuentran ampliamente desarrolladas en la segunda parte de la tesis.

Siendo la Teoría de Grafos la base teórica fundamental para la comprensión y posterior análisis de la organización de las redes sociales, el RAT se hace imprescindible dentro de un entorno de enormes volúmenes de información “*Big Data*”, para poder identificar aquellos textos de interés y posteriormente resumir y analizar su contenido tanto desde su estructura gramatical como del análisis de los sentimientos, que llamaremos “*Sentiment Analysis*”, que los usuarios han reflejado a través de sus conversaciones en los medios sociales.

Respecto a la Teoría de Grafos, hemos revisado históricamente su evolución y perfeccionamiento partiendo de sus orígenes y estudiando los autores con aportaciones más relevantes. Para ello, contaremos con aportaciones desde campos como la sociología, la psicología, las matemáticas, la antropología, el marketing y la comunicación.

Superada esta fase teórica de la tesis, avanzaremos en el estudio empírico trabajando con la técnica de la netnografía, cuyas bases se asientan en la investigación cualitativa y parte de la antropología cultural. Su voluntad manifiesta es desarrollar un estudio del contexto en el que tiene lugar la conducta del consumidor en las comunidades virtuales.

Las principales técnicas de campo etnográficas son la observación (de forma participativa o no), la conversación, la entrevista en profundidad, el análisis de las redes sociales, el método genealógico, las llamadas historias de vida y el análisis documental. (Del Fresno, 2011:54)^[3]

En la actualidad, gracias a los medios tecnológicos disponibles, es habitual que los etnógrafos sociales hagan los estudios mientras el acontecimiento todavía está operativo, es decir, mientras está sucediendo. La duración del período de obtención de los datos también acostumbra a ser más breve, en parte porque el etnógrafo puede obtener y guardar digitalmente grandes cantidades de información de forma rápida y porque la orientación de los estudios acostumbra a ser más específica, a escala que podríamos denominar micro. (Hammersley, 2006:4-5)^[4]

La netnografía, a diferencia de la etnografía, permite que los efectos sobre la realidad social que quiere describir sean inexistentes. De este modo, presenta posibilidades naturalistas: “permite la observación no participante si es oportuno y ético para el objeto de estudio” (Del Fresno, 2011:55)^[3] (Nelson, 2005:90)^[5]

Robert V. Kozinets¹ presentó el año 1995 y por primera vez en público la netnografía como técnica de investigación. Se considera una especialización de la etnografía para ser gestionada a través de un ordenador y con ello poder analizar las características de los entornos sociales que se presentan en la red. Como todo estudio social, su finalidad es obtener información relevante para comprender mejor determinados aspectos de las comunidades humanas y las singularidades existentes dentro de ellas. (Kozinets, 2010)^[6]

A medida que nuestras relaciones sociales se han ido incorporando a lo digital, los sociólogos de todo el mundo han comprendido certeramente que el comportamiento de las personas en Internet es esencial para comprender la sociedad contemporánea. Para ello, es necesario monitorizar y realizar un seguimiento de sus acciones a través de programas y sistemas de análisis computarizados. Al mismo tiempo, por motivos análogos al los de la sociología, la netnografía se utiliza, por ejemplo, en la investigación de mercados, el análisis del consumidor, el marketing, la publicidad, el aprendizaje o la creatividad.

De hecho, en el mundo científico, la conclusión de que es imprescindible incorporar Internet en los estudios sociales y culturales si se quiere alcanzar un resultado satisfactorio está cada vez más extendida. Esta conclusión va asociada directamente a la necesidad de utilizar los programas y aplicaciones informáticas que permiten trabajar con los datos obtenidos en el mundo digital. (Kozinets, 2010:2)^[6]

Según Kozinets, estas son las etapas a seguir para aplicar el método netnográfico:

- Planificación de la investigación
- Aproximación cultural al objeto de estudio
- Observación y análisis de la información

¹ Robert V. Kozinets. Creador de la técnica de investigación: “Netnografía” en 1995. Reconocido investigador en Marketing y Redes Sociales.

- Interpretación de los resultados
- Revisión del cumplimiento de las consideraciones éticas
- Presentación de los resultados a los integrantes del grupo social estudiado y obtención de la retroalimentación por parte de éstos.

Otra característica propia de la netnografía es que el trabajo de campo no se desarrolla en una localización física sino que lo que entendemos por “campo” es aquí el espacio social común señalado por el código pro-social que facilita las interrelaciones. (Del Fresno, 2011:84)^[3]

Como hemos apuntado anteriormente, una de las ventajas de la netnografía es el hecho de poder observar las interacciones en su propio contexto y de forma no intrusiva, a diferencia de las técnicas como el *Focus Group* o las entrevistas en profundidad. Así, la información que se obtiene a través de este método es natural y espontáneo. Tal y como comenta el propio Kozinets: “This data is raw, It is authentic. Spontaneous. Indigenous. Unforced. Unadorned. Powerful. Highly involved and often spectacularly creative” (Kozinets, 2010:2)^[6]

Otro elemento interesante a destacar es que el investigador tiene la información a su disposición las 24 horas del día y todos los días, sin excepción. Por lo tanto, puede regresar a la fuente de información en cualquier momento. (De Valck, 2009:197)^[7]

En paralelo a la netnografía, la folcsonomía es una herramienta que también hemos utilizado dentro del campo de la computación. Etimológicamente puede valer nos su significado. Su procedencia radica en la “taxonomía”, del griego “taxis” y “somos”, cuyo significado es: clasificación, ordenación y gestión. En cuanto a “folc”, su procedencia es alemana y significa “pueblo”. Etimológicamente, folcsonomía expresa literalmente “clasificación gestionada por el pueblo”^[8]

La folcsonomía proviene de un concepto creado por Thomas Van de Wal y es el resultado del etiquetado realizado de forma libre por parte de los usuarios, de información y objetos (cualquier cosa con una URL) para su propio uso. El etiquetado se desarrolla en un entorno

social (normalmente compartido y abierto a los demás). La folcsonomía se crea a través del acto de etiquetado que hace el usuario que consume información.

En su propia web, el mismo Thomas Van der Wal explica el origen del término folcsonomía (*folksonomy*) de la siguiente manera:

On July 23, 2004 in the IA Institute (then called the Asylomar Institute for Information Architecture (AIFIA)) closed list serve Gene Smith asked, “Some of you might have noticed services like Furl, Flickr and Del.icio.us using user-defined labels or tags to organize and share information.... Is there a name for this kind of informal social classification?”. After a few other people answered some other related questions Eric Scheid of Ironclad Information Architecture responded with “folk classification”. On July 24, 2004 I responded just after that with, “So the user-created bottom-up categorical structure development with an emergent thesaurus would become a Folksonomy? (Van der Wal, 2007)^[9]

En cuanto a la definición, estas son sus palabras:

Folksonomy is the result of personal free tagging of information and objects (anything with a URL) for one’s own retrieval. The tagging is done in a social environment (usually shared and open to others). Folksonomy is created from the act of tagging by the person consuming the information.

The value in this external tagging is derived from people using their own vocabulary and adding explicit meaning, which may come from inferred understanding of the information/object. People are not so much categorizing, as providing a means to connect items (placing hooks) to provide their meaning in their own understanding.

Comentaremos ahora los aspectos más relevantes de los textos publicados en su web: “etiquetado libre y personal”, “etiquetado realizado en un entorno social”, “inclusión de un significado explícito”. Con todo ello, entendemos aquí que los individuos, a través del etiquetado de las palabras clave, la selección de los dominios de los sitios web, el uso de los emoticonos y también del uso de sus cuentas (personales o empresariales) en las redes sociales, aportan una información relevante sobre sus gustos, preferencias, hábitos o temas de interés que completan la información y el conocimiento que se puede obtener de ellos.

Del mismo modo que hemos presentado las ventajas del uso del método etnográfico, la folcsonomía presenta varias ventajas que repercutirán positivamente en la calidad de los resultados de nuestra investigación. Lo mostramos en la siguiente enumeración:

- Dado a su funcionamiento sencillo, permite reducir el tiempo de clasificación.
- Al ser etiquetas o términos elegidos con libertad por los propios usuarios, las folcsonomías asocian las verdaderas necesidades de los usuarios con sus formas habituales de expresión. En nuestro estudio hemos utilizado las palabras clave que los usuarios utilizan habitualmente para identificar o buscar información.
- Permiten identificar nuevas áreas de interés de una pequeña parte de la población. Por ejemplo, la herramienta llamada “Machine Learning” incorpora en el estudio nuevas tendencias que aparecen en las conversaciones de los usuarios y que en nuestro caso se atribuyen a los *hashtags* más utilizados por los usuarios.

Es preponderante señalar que Twitter utiliza este sistema de etiquetado para gestionar, categorizar y resumir ciertos tipos de información: permite el uso de emoticonos, palabras clave y “*hashtags*”, entre otras características que hemos tomado como datos para nuestra investigación. Gracias a ello, nos ha sido posible categorizar a los usuarios de Twitter estudiados en función de las etiquetas que han usado para sus contenidos. Al mismo tiempo, hemos logrado resumir los contenidos que son más relevantes a nivel de la expresión de los sentimientos.

La decisión de haber aplicado un criterio interdisciplinar en la elección de los métodos de investigación ha sido la expresa voluntad de evitar situarnos en una posición reduccionista y que no captase la complejidad y riqueza que ofrece, por contra, la combinación de los métodos elegidos. Para Busquet, Medina y Sort, tener una pluralidad de métodos a la hora de emprender una investigación tiene rasgos positivos:

La primera virtud de un investigador social es la curiosidad que le empuja a ir más allá de su campo específico. Esta actitud abierta significa poner en cuestión las fronteras convencionales y arbitrarias que separan las actuales disciplinas del conocimiento social. (Busquet, 2006:33)^[10]

En este sentido, la conveniencia de utilizar diversos métodos o técnicas en un mismo trabajo de investigación no se plantea en profundidad hasta la segunda mitad del siglo XX, momento en el que se empieza a debatir la adecuación de combinar la encuesta cualitativa con el trabajo cuantitativo de campo o la entrevista con la observación participante. (Cea, 1998:47-48)^[11]

Consideramos que las herramientas actuales desarrolladas con las evolución en el entorno digital no han hecho sino aumentar esa necesidad y pertinencia del uso de métodos de varias disciplinas en la investigación.

Uno de los partidarios de esta aproximación multimétodo es Sieber, que describe sus ventajas de la manera siguiente:

The integration of research techniques within a single project opens up enormous opportunities for mutual advantages in each of three major phases-design, data collection, and analysis. These mutual benefits are not merely quantitative (although obviously more information can be gathered by a combination of techniques) But qualitative as well, one could almost say that a new style of research is born of the marriage of survey and fieldwork methodologies. (Sieber, 1973: 337)^[12]

En esta tesis suscribimos dicha visión y consideramos que es la única que nos permitirá afrontar el trabajo con las suficientes herramientas como para lograr un resultado satisfactorio.

TERMINOLOGÍA

Terminología

El desarrollo de las nuevas tecnologías y la aparición de productos como las *tablets* o los *smartphones* ha demandado la creación de numerosos términos que definan las nuevas prestaciones y posibilidades que ofrecen. Al mismo tiempo, Internet y la *World Wide Web* generan también nuevos productos, diseños y aplicaciones que, a su vez, requieren su propia terminología específica. En nuestro caso, al centrarnos en una investigación a través de Twitter, precisaremos definir algunos conceptos que serán clave para el desarrollo de la tesis. Varios de ellos se aplican, además, en el ámbito profesional. Por otro lado, en cuanto a la metodología, será pertinente definir varios conceptos de igual relevancia.

Análisis de Redes Sociales (*Social Network Analysis*):

El Análisis de Redes Sociales consiste en el estudio metódico (utilizando un determinado método) de las interacciones que tienen lugar en las redes sociales entre sus usuarios.

Social network analysis (SNA) is the methodical analysis of social networks. Social network analysis views social relationships in terms of network theory, consisting of nodes (representing individual actors within the network) and ties (which represent relationships between the individuals, such as friendship, kinship, organizational position, sexual relationships, etc.). These networks are often depicted in a social network diagram, where nodes are represented as points and ties are represented as lines.^[13]

Álgebra de Boole (álgebra booleana):

El álgebra de Boole, cuyo nombre honora a su fundador George Boole (Inglaterra 1815 - Irlanda 1864), es una ramificación algebraica que el matemático inglés consideró parte de un sistema lógico. El trabajo de Boole fue un intento de utilizar las técnicas algebraicas para tratar expresiones de la lógica proposicional: *Las interpretaciones respectivas de los símbolos 0 y 1 en el sistema de lógica son Nada y Universo.*^[14] Dicha estructura algebraica

esquematiza las operaciones lógicas como Y (conjunción), O (disyunción), NO (negación) y SI (afirmación).

En la actualidad, el álgebra de Boole se aplica de forma generalizada en el ámbito del diseño electrónico. El ingeniero electrónico y matemático Claude Shannon fue el primero en aplicarla en el diseño de circuitos de conmutación eléctrica biestables, en 1948. Esta lógica se puede aplicar a dos campos:

- Al análisis, porque es una forma concreta de describir cómo funcionan los circuitos.
- Al diseño, ya que teniendo una función aplicamos dicha álgebra, para poder desarrollar una implementación de la función.

Capital social:

En redes sociales, realizando un acercamiento general, se entiende por capital social el valor y la calidad de las relaciones establecidas entre los usuarios que hacen uso de ellas, ya sean entre personas, empresas, instituciones o entre unos y otros. Este concepto de capital social parte de una visión instrumental y endógena que solo tiene en cuenta las relaciones dentro de las mismas redes y, como resultado, las investigaciones realizadas al respecto permiten segmentar los vínculos (relaciones) entre usuarios de forma cualitativa. Por ejemplo, pueden catalogarse como “fuertes”, “débiles”, “relevantes”, etc. Análogamente, se manejan variables que situarían esos vínculos en categorías asociadas a relaciones muy cercanas e indispensables (“vitales”) o, por ejemplo, en relaciones óptimas para el desarrollo profesional. El término sigue siendo motivo de estudio y discusión actualmente y hay diversas concepciones que pueden llegar a ser contrapuestos. (Prell, 2012:62)^[15]

Ego Network:

Es una vía alternativa para estudiar el uso de las redes sociales centrada en el individuo y en su relación con los otros. Los informes y datos elaborados con esta vía se basan únicamente en la visión que los usuarios tienen de sí mismos (Wellman, 1983)^[16]. Con las *Ego Network*, cada usuario es considerado el centro de su propia red. Al usuario estudiado se le denomina “yo” y, a todas sus relaciones y al resto de los usuarios, “otros”.

Enumeramos tres puntos esenciales de las *Ego Network* (Prell, 2012:118)^[17]:

- Un ego es nodo individual con capacidad de actuar. Como el propio nombre indica, el centro de una *Ego Network* es el ego.
- Las *Ego Network* se componen de un sujeto, el “ego”, y de sus contactos inmediatos, conocidos como “otros”.
- Un ego, los otros y los vínculos que establecen entre sí conforman la *Ego Network* y son una gran parte del objeto de estudio.

Folcsonomía (*Folksonomy*):

La Folcsonomía es el resultado del etiquetaje manual de información, objetos, cualquier cosa con una dirección de Internet, para el almacenaje o archivo propios.

Folksonomy is the result of personal free tagging of information and objects (anything with a URL) for one’s own retrieval. The tagging is done in a social environment (usually shared and open to others). Folksonomy is created from the act of tagging by the person consuming the information.

The value in this external tagging is derived from people using their own vocabulary and adding explicit meaning, which may come from inferred understanding of the information/object. People are not so much categorizing, as providing a means to connect items (placing hooks) to provide their meaning in their own understanding.

In a few conversations around folksonomy and tagging in 2004 I stated, “folksonomy is tagging that works”. This is still a strong belief the three tenets of a folksonomy: 1) tag; 2) object being tagged; and 3) identity, are core to disambiguation of tag terms and provide for a rich understanding of the object being tagged. (Van der Wal, 2007)^[9]

La Fuerza de los lazos débiles (*The Strength of weak ties*)

Teoría formulada por Mark Granovetter² y publicada en 1973 en *The American Journal of Sociology* que concluye, tras los estudios realizados por el mismo Granovetter, que los lazos débiles (con poco o ningún contacto entre sí y sin un vínculo emocional) son mejores y más relevantes para obtener oportunidades de encontrar trabajo que los lazos fuertes (familiares, amigos y otros lazos de contacto regular y con vínculo emocional). (Prell, 2012:76)^[17]

***Influencer* (“con capacidad de influencia”):**

Persona con notable y exitosa exposición pública, principalmente difundida y fomentada por ella misma a través de sus redes sociales, que cuenta con una comunidad elevada de seguidores caracterizada por prestar especial atención a los contenidos que comparte y una predisposición a copiarlos o emularlos. Podemos encontrar muchas tipologías de *influencers*, desde aquellos que son conocidos por tener un criterio valioso en una determinada serie de productos o servicios concretos a los que son admirados por el estilo de vida. El seguimiento que logran suele estar basado en la credibilidad, la confianza y la empatía que despiertan entre su comunidad. Sus opiniones pueden, por lo tanto, repercutir positiva o negativamente en el consumo de determinados productos y servicios, motivo por el cual muchas empresas intentan (ya sea con el suministro gratuito de productos o con el patrocinio económico) que los *influencers* hablen, mencionen o compartan contenido (textual, visual o audiovisual) recomendando favorablemente su marca. Con la expansión y crecimiento del fenómeno de los *influencers* se dan numerosos casos en los que esto logran vivir únicamente de aquello que recomiendan y por lo que son reconocidos.

***Latent Semantic Analysis (LSA)* o *Latent Semantic Indexing (LSI)*:**

Método de análisis semántico que se basa en la suposición de que existe una estructura latente tal en el uso de las palabras que hará que las palabras con significado parecido, como

² Mark Granovetter (EEUU, 1943). Sociólogo y profesor de la Universidad de Standford. Elaboró la teoría de “La fuerza de los lazos débiles”.

por ejemplo sinónimos, puedan aparecer en espacios parecidos del texto. Se considera una de las técnicas recientes que más utilizadas y que mayor interés han despertado. (Deerwester et al., 1988)^[18]

El LSA, cuando analiza el Procesamiento Natural del Lenguaje, establece relaciones entre grupos de textos, documentos y los términos que estos contienen. Expresa el contenido contando las palabras por párrafos (las líneas representan palabras únicas y las columnas representan cada párrafo) y, junto con la técnica matemática llamada “*Singular Value Decomposition*” (SVD), sirve para reducir el número de columnas manteniendo una estructura similar entre líneas. Las palabras se comparan a partir del coseno del ángulo entre dos vectores formado por cada dos líneas. Valores cercanos a 1 representan palabras muy similares, mientras que valores similares a 0 representan palabras muy distintas. En cuanto a las posibles aplicaciones de la LSA encontramos, por ejemplo, que es de gran utilidad para la recuperación de información (*Information Retrieval*), clasificación y agrupación de grandes cantidades de documentos de texto. En otra de sus múltiples aplicaciones, el LSA permite hallar documentos similares en dos idiomas distintos habiendo analizado previamente otros textos de ambos idiomas.

Linguistic Inquiry and Word Count (LIWC):

El LIWC^[19] es un programa de análisis de textos diseñado por James W. Pennebaker, Roger J. Booth y Martha E. Francis que calcula el uso que realizan las personas de diferentes categorías de palabras desde cualquier fuente de texto incluyendo correos electrónicos, discursos, artículos, o la transcripción de conversaciones cotidianas. De forma inmediata, analiza los contenidos de cualquier texto determinando el uso de emociones positivas o negativas, auto-referencias, palabras causales y hasta 82 otras dimensiones lingüísticas.

El programa puede analizar cientos de archivos de texto ASCII estándar o documentos de Microsoft Word en cuestión de segundos. La versión “LIWC2007” del programa permite también crear diccionarios propios y con ello es posible mejorar la precisión del análisis centrándolo solo en los términos relevantes para los objetivos del estudio. El desarrollo del programa LIWC se ha creado con el fin de proporcionar un método eficiente y eficaz para el

estudio de los distintos componentes de los distintos procesos estructurales presentes en las manifestaciones verbales y escritas de los individuos, sean emocionales o cognitivas.

Este método se sustenta en la evidencia científica que muestra que es posible obtener vías de acceso a los mundos emocionales y cognitivos de una persona en función de la manera que tiene de expresarse verbalmente y por escrito. Además, durante las últimas tres décadas, los investigadores han proporcionado pruebas que sugieren que la salud física y mental de las personas se puede predecir por las palabras que usan (Gottschalk y Glasser, 1969)^[20] (Rosenberg y Tucker, 1978)^[21] (Stiles, 1992)^[22]. Más recientemente, un gran número de estudios han encontrado que las personas capaces de escribir o hablar acerca de experiencias con alta carga emocional presentan mejoras tanto física como mentalmente (Pennebaker, 1997)^[23] (Smyth, 1997)^[24]. Los análisis de texto basados en estos estudios indican que las personas a las que la escritura les produce mayores beneficios tienden a utilizar tasas relativamente altas de palabras emocionales positivas, un número moderado de las palabras que expresan emociones negativas, y, lo más importante, utilizan de principio a fin un número creciente de palabras que revelan sus procesos cognitivos (Pennebaker y Francis, 1996)^[25] (Pennebaker, Mayne y Francis, 1997)^[26]

La primera aplicación del programa LIWC fue desarrollada como parte de un estudio exploratorio de la lengua y la divulgación (Pennebaker y Francis, 1993)^[27]. Como hemos apuntado, la segunda versión, LIWC2007, es una revisión actualizada de la aplicación original y ha sido diseñada con capacidad para analizar palabra por palabra el texto escrito. Tras el análisis, LIWC2007 calcula el porcentaje de las palabras del texto que coinciden con cada una de las dimensiones del lenguaje elegidas previamente (como señalamos, pueden ser hasta 82) y genera un informe legible como texto delimitado por tabuladores y compatible con programas como el SPSS para Windows o el Microsoft Office Excel.

Machine Learning (ML):

El *Machine Learning* (ML) o Aprendizaje automático (Schapire, 2008)^[28] estudia los algoritmos informáticos para generar un aprendizaje que permita a los computadores aprender a realizar ciertas acciones o procesos por sí mismos. El aprendizaje que realizan las

computadoras se basa siempre en observaciones o datos que pueden provenir tanto de ejemplos similares a los que la computadora debe aprender, experiencias directas o cierto tipo de tests. En este sentido, el ML consiste en que la propia computadora aprenda a hacer algo mejor en un futuro a base de sus experiencias pasadas con lo que se le ha mandado hacer, y todo ello con métodos automáticos y sin intervención humana. Un caso típico del ML es la programación de los filtros de spam, basado en el paradigma que podríamos llamar “aprender a base de ejemplos”. En vez de programar directamente el ordenador para que tenga un filtro de spam, buscamos que el propio programa encuentre su propio filtro basándose en los ejemplos de correo spam que le hemos ido enseñado. Por todo ello, se considera el Machine Learning como una de las grandes áreas dentro del desarrollo de la Inteligencia Artificial, ya que no se considerará inteligente nada que no pueda ser capaz de aprender por sí mismo. Además del campo de la Inteligencia Artificial, el ML abarca otros campos del saber como la estadística, la matemática, la física o la ciencia computacional teórica, entre otros.

Massive Online Analysis (MOA):

El MOA es el recurso de código abierto más popular para analizar contenido compartido en internet en tiempo real.

MOA is the most popular open source framework for data stream mining. It includes a collection of machine learning algorithms (classification, regression, clustering, outlier detection, concept drift detection and recommender systems) and tools for evaluation.

MOA performs Big Data stream mining in real time, and large scale Machine Learning. MOA can be extended with new mining algorithms, and new stream generators or evaluation measures. The goal is to provide a Benchmark suite for the stream mining community.^[29]

Netnografía:

La netnografía se considera un método de investigación derivado de las técnicas etnográficas desarrolladas por la investigación antropológica pero trasladadas al entorno de la *World Wide Web*. Gracias a las nuevas posibilidades de estudio del comportamiento de los usuarios a

escala masiva que permiten las herramientas surgidas con el desarrollo de la programación informática, pueden filtrarse grandes cantidades de datos cuyo resultado puede revelar valiosas informaciones acerca de las personas y las comunidades a las que pertenecen. (Kozinets, 2002)^[30]

Opinion Mining:

El *Opinion Mining* (Pang y Lee, 2008)^[31], también llamado *Sentiment Analysis*, es el campo de estudio que analiza las opiniones, sentimientos, valoraciones, apreciaciones, actitudes y emociones de las personas respecto a entidades tales como productos, servicios, organizaciones, individuos, temas, eventos, tópicos y sus atributos. En el siguiente apartado de la tesis se desarrollará ampliamente este concepto con toda su aplicación práctica.

PageRank:

El *PageRank*^[32] permite consultar la importancia de una página según los cálculos que realiza Google a partir de determinados factores como, por ejemplo, el número de enlaces que conducen a dicha página. Las páginas web con un *PageRank* más alto tienen más posibilidades de aparecer al principio de los resultados de búsqueda de Google.

Procesamiento del Lenguaje Natural, PLN (*Natural language processing, NLP*):

El PLN es el método simbólico más directo y común utilizado para acercar el lenguaje computacional (usado por ordenadores) y el lenguaje natural (utilizado por las personas). Su utilización permite procesar grandes datos de información almacenada en las computadoras y, a través de su estudio con diversas herramientas, obtener datos relevantes y útiles para multitud de fines.

Generally, natural language (used by humans) is the most direct method and the symbol system most in use to express human ideas and pass on information. There is a gap between formal languages (used by a computer) and natural languages. Communication between computers and humans is only possible when much research is aimed at bridging this gap.

The approach to bridge the gap is often named natural language processing (NLP), natural language understanding, or computational linguistics. (Allen, 1987)^[33] (Harris, 1985)^[34]

Prosumer:

Concepto acuñado por Alvin Toffler^[35] en el año 1980 que unifica los conceptos “consumidor” y productor” en el sentido de “considerar a los consumidores también productores”. Toffler adelantó el paso siguiente en el mercado de bienes y servicios tras la producción en masa e industrial. Según preveía, las empresas cada vez tenderían a “personalizar” los productos en busca de una mayor identificación con las singularidades de cada consumidor. Al mismo tiempo, como proceso paralelo, los consumidores pasarían a identificarse con ciertos bienes y servicios (desde una marca de ropa hasta un evento cultural), contribuyendo en su difusión, valoración, etc. La definición sería notablemente acogida por el ámbito académico y recibiría numerosas contribuciones y matices, como la relevante aportación de George Ritzer.^[36]

Redes Bayesianas (*Bayesian Networks*):

Las Redes Bayesianas son un método gráfico específico.

Bayesian networks are a specific type of graphical model: a directed acyclic graph (Neapolitan, 1989)^[37]. That is, all of the edges in the graph are directed (i.e., they point in a particular direction) and there are no cycles (i.e., there is no way to start from any node and travel along a set of directed edges in the correct direction and arrive back at the starting node). (Stephenson, 2000)^[38]

Small Worlds (El problema del pequeño mundo):

La manera más simple de formular *El problema del pequeño mundo* es con la pregunta “¿Cuál es la probabilidad de que dos personas, seleccionadas arbitrariamente de entre una amplia población, como los Estados Unidos, se conozcan entre sí?”.

The simplest way of formulating the small world problem is “what is the probability that any two people, selected arbitrarily from a large population, such as the United States, will know each other?” A more interesting formulation, however, takes account of the fact that, while persons a and z may not know each other directly, they may share one or more mutual acquaintances; that is, there may exist a set of individuals, B , (consisting of individuals ($b_1, b_2... b_n$) who know both a and z and thus link them to one another. More generally, a and z may be connected not by any single common acquaintance, but by a series of such intermediaries, $a-b-c... -y-z$; i.e., a knows b (and no one else in the chain); b knows a and in addition knows c , c in turn knows d , etc. (Travers, 1969)^[39]

Social Media Analytics:

El *Social Media Analytics* emerge de la confluencia del *Social Network Analysis*, el *Machine Learning*, el *Data Mining*, el *Information Retrieval* (IR y el NLP). La unión de este conjunto de herramientas y métodos permite realizar estudios del contenido que los usuarios comparten en las redes sociales obteniendo resultados que no podrían alcanzarse de ningún modo usando uno de los métodos de forma aislada.

This has given rise to the emerging discipline of Social Media Analytics, which draws from Social Network Analysis, Machine Learning, Data Mining, Information Retrieval (IR), and Natural Language Processing (NLP). (Melville, 2009)^[40]

The perspective of social network analysis—which comprises both method and theory—mitigates against studying any single relationship in isolation from the network of which it is

part. This is because the dyad, or relationship between two actors, is the building block of a network, but is itself conditioned by the network.

Across social network studies, the actors or the nodes have been variously defined as individuals, groups, companies, or even countries. The relationship or tie is a flow of resources that can be material or non-material (Wasserman y Faust, 1999:4)^[41]. The resources might include social support, emotional support, companionship, time, information, expertise, money, business transactions, shared activity, and so on.

Socialnomics:

Socialnomics es el valor que se genera a través del contenido creado y compartido en las redes sociales y que tiene una influencia real en diversos campos (por ejemplo, en el campo económico o político).

“Socialnomics is the value created and shared via social media and its efficient influence on outcomes (economic, political, relational, etc,...). Or, more simply put, it’s word of mouth on digital steroids. A subset of this is that in the future we will no longer search for products and services; rather, they will find us via social media” (Qualman, 2013:XVII)^[42]

Sociometría (Sociometry):

En una publicación de 1937, Jacob L. Moreno³ fundó la “Sociometría” (*Sociometry*) para la medición de las relaciones interpersonales en grupos pequeños, e inventó el concepto “*Sociogram*” con el fin de estudiar las relaciones interpersonales como si fuesen estructuras, en las cuales las personas, o cualquier unidad social, están representadas por puntos en un espacio bidimensional. En este espacio las relaciones entre estas unidades se representan a través de líneas que enlazan los puntos correspondientes.

³ Jacob L. Moreno (1.889 – 1.974) Nació en Rumanía y desarrolló su profesión EEUU. Psiquiatra y fundador de la sociometría.

Teoría de las Redes Sociales (*Social Network Theory*):

La *Teoría de las Redes Sociales* se define en contraposición a la teoría que define a la sociedad como una mera suma de individuos. En vez de ello, estudia las relaciones entre los individuos y en cómo dichas relaciones modelan la sociedad a través de redes o lazos que denomina “nodos”.

Social network theory contrasts with the type of sociological theory that defines society as built up of individuals. It starts instead from the relations between individuals, and models society as constituted of networks made up of sets of the relations or ties between the nodes. Wasserman and Faust^[41] identify four additional fundamental principles of models built using social network theory: independence of actors; relations or ties consisting in the flow or transfer of resources; the constraining and/or enabling of individual actors by networks; and the generation of long-lasting ties and networks by social structures.

Teoría de los Seis Grados:

Este concepto está también desarrollado ampliamente en los antecedentes a la Teoría de Grafos del siguiente capítulo de este trabajo.

The theory of six degrees of separation states that any two random-selected people on this world can get to know each other by no more than six steps of intermediate friend chains (Zhang y Wanqing, 2009)^[43]. It was originally set out by Frigyes Karinthy⁴

Con la aparición de Internet y las nuevas tecnologías que permiten aumentar exponencialmente las conexiones virtuales entre individuos, la Teoría de los Seis Grados regresará al interés académico y autores como Duncan Watts la actualizarán estudiando las variaciones en el los grados de separación que ha generado el surgimiento de Internet. (Watts, 2004)^[44]

⁴ Frigyes Karinthy. (Budapest 1887 - Siófok 1938). Escritor húngaro. f en 1929

Text Mining, TM (Minería de textos):

El *Text Mining* es el análisis de datos contenido en el lenguaje natural escrito.

Text mining is the analysis of data contained in natural language text. Text mining works by transposing words and phrases in unstructured data into numerical values which can then be linked with structured data in a database and analyzed with traditional data mining techniques.

Text mining is the analysis of data contained in natural language text. The application of text mining techniques to solve business problems is called text analytics.

Text mining can help an organization derive potentially valuable business insights from text-based content such as word documents, email and postings on social media streams like Facebook, Twitter and LinkedIn. Mining unstructured data with natural language processing (NLP), statistical modeling and machine learning techniques can be challenging, however, because natural language text is often inconsistent. It contains ambiguities caused by inconsistent syntax and semantics, including slang, language specific to vertical industries and age groups, double entendres and sarcasm.

Text analytics software can help by transposing words and phrases in unstructured data into numerical values which can then be linked with structured data in a database and analyzed with traditional data mining techniques. With an iterative approach, an organization can successfully use text analytics to gain insight into content-specific values such as sentiment, emotion, intensity and relevance. Because text analytics technology is still considered to be an emerging technology, however, results and depth of analysis can vary wildly from vendor to vendor. (Rouse, 2015)^[45]

Text summarization:

El *Text summarization* permite obtener de manera automática y a través del análisis por computadora del contenido de un determinado texto, los puntos más relevantes del mismo.

Also know as Automatic summarization: systems designed to take a single article, a cluster of news articles, a broadcast news show, or an email thread as input, and produce a concise and fluent summary of the most important information. (Nenkova y McKeown, 2011)^[46]

Web Semántica:

La Web Semántica proporciona un marco común que permite que los datos sean compartidos y reutilizados a través de aplicaciones, empresas y fronteras comunitarias. Es un esfuerzo colaborativo liderado por el W3C con la participación de un gran número de investigadores y socios industriales. Está basado en *Resource Description Framework* (RDF) e integra una variedad de aplicaciones utilizando XML para la sintaxis y URL para las denominaciones (www.w3.org/2001/sw/). (Codina, 2006)^[47]

La Web Semántica (*Semantic Web*) pretende facilitar a los usuarios el acceso, la búsqueda y selección de información de acuerdo con sus intereses, de una forma sencilla y eficaz y con los mejores parámetros de calidad. (Berners-Lee et al., 2001)^[48] (Bravo y Redondo, 2004)^[49]

ESTRUCTURA Y PARTES DEL TRABAJO DE INVESTIGACIÓN

Estructura y partes del trabajo de investigación

La estructura de esta tesis está organizada en seis capítulos principales que a su vez se subdividen en secciones. Explicaremos resumidamente el contenido y la finalidad de cada una de las secciones contenidas en los diversos capítulos.

Con la fijación de los objetivos podemos determinar cómo realizaremos el trabajo de campo y qué herramientas son las más apropiadas para lograr alcanzarlos. En nuestro caso, el objetivo es demostrar que las conversaciones mantenidas en las redes sociales por sus usuarios nos aportan información de alto valor que, si se estudia y analiza con métodos científicos, pueden servir para realizar *insights* publicitarios. A este objetivo principal se incorporan otros dos, también de gran importancia, relacionados con el primero. El segundo objetivo es demostrar y evidenciar que los resultados obtenidos con nuestro estudio podrá ser válido más allá del departamento de publicidad o de una empresa de publicidad y que, por lo tanto será útil en la toma de decisiones de otros departamentos de las empresas analizadas, como puede ser el mismo equipo directivo o, por ejemplo, el área responsable del diseño de producto. Como tercer y último gran objetivo queremos evidenciar la relevancia de la variable *Sentiment* en las comunicaciones de los usuarios y de su influencia en el curso de las conversaciones. La variable *Sentiment*, que engloba el conjunto de emociones y estados de ánimo que expresan los usuarios en los mensajes, es relevante para desarrollar los *insights* y su estudio puede ser útil para los otros departamentos, contribuyendo a alcanzar nuestro segundo objetivo.

Las tres hipótesis que planteamos guardan una relación directa con cada uno de los objetivos. En primer lugar, consideraremos que los usuarios reflejan en las redes sociales sus pensamientos, emociones y opiniones con una intención comunicativa que coincide, en la inmensa mayoría de los casos, con aquello que el usuario expresaría en otros ámbitos de su vida fuera de las redes sociales. Las conversaciones, expresiones o decisiones que tomaría en un ámbito familiar, laboral o en su círculo de amistades guardan una relación directa con la que mantiene en los mismos círculos pero en un entorno digital. A esto añadimos que la comunicación que el usuario mantiene ahora con las empresas y las marcas abre nuevos

caminos y posibilidades de interacción que el usuario aprovecha, en mayor o menor medida, con muy diversas finalidades. En un segundo lugar, situamos como hipótesis que las empresas. Nuestra segunda hipótesis sostiene que en el ámbito empresarial ya se considera el trabajo realizado en las redes sociales como una de las prioridades del área de comunicación y los esfuerzos destinados al cuidado de la imagen marca y todo aquello que aspira a transmitir y representar. Fruto de este interés, y gracias a esta hipótesis, podremos ofrecer nuevos avances que representamos en los objetivos mencionados. Como tercera hipótesis sostenemos que las emociones y los estados de ánimo cobran una importancia determinante en el curso y la evolución de las conversaciones en la red. En consonancia con ello, consideramos también que los usuarios pueden expresarlas de forma satisfactoria usando los códigos disponibles en las redes sociales (*hashtags*, emoticonos, etc.). Si este planteamiento resulta correcto, el tercer objetivo que hemos fijado y que aspira a convertir la detección de la variable *Sentiment* en información útil estará en condiciones óptimas para poder alcanzarlo.

PRIMERA PARTE

1.- MARCO TEÓRICO Y CONCEPTUAL

1. Marco teórico y conceptual

1.1. La comunicación en los medios sociales

Para desarrollar la importancia capital de los medios sociales y de su penetración en la cotidianidad de miles de millones de personas alrededor del mundo, puede ser de utilidad realizar un acercamiento desde los medios de comunicación tradicionales, ya sea la prensa escrita, la televisión o la radio. Si en la introducción de la primera parte de esta tesis hablábamos de la importancia que han tomado las redes sociales como medio de información para sus usuarios, nos centraremos en esta introducción en la relevancia de los propios medios sociales como apertura de posibilidades comunicativas que plantea el poder acceder a su uso.

Es especialmente representativo que la revista *Time* dedicase su portada anual “*Person of the year*” a “*You*”^[50], en el año 2006. La portada, que muestra un ordenador con la pantalla reflectante y la palabra “*You*” escrita en ella, viene a decir que “usted”, como usuario de la red, es el gran protagonista del año por la importancia que adquieren su comportamiento y creatividad en el manejo de la información a través del uso de la red y sus herramientas sociales. Se mencionan, entre otros, Wikipedia, YouTube, Myspace y Facebook, cuyos contenidos se generan íntegramente por sus usuarios, sin que haya un medio de comunicación de estructura vertical que seleccione y filtre lo que debe publicarse y lo que no. Esta portada podría verse como una continuación, más de treinta años después, de una portada de la misma revista en el año 1982. Ese año, se eligió “*Machine of the year*” a los “*Personal computers*”^[51]. Fue en la década de los ochenta que tuvo lugar el cambio de paradigma en el que los ordenadores dejaron de ser una herramienta empresarial y científica para convertirse en “máquinas de uso personal” al alcance de un público masivo. Con la llegada de Internet y, en concreto, de las posibilidades de la web 2.0, cuyas características trataremos más adelante, comprenderemos el por qué *Time* elegiría la mencionada portada de 2006. En la misma línea, el medio de comunicación *ABC News* nombraría a los *bloggers* “*People of the year*”^[52] en el año 2004, por la creciente influencia de sus opiniones en la esfera pública y en los campos de especialización que algunos de ellos iban ocupando. La figura de los *influencers*, término que

hemos definido en la terminología de la primera parte de esta tesis, se deriva de la toma de conciencia de que la opinión de los *bloggers* (y, posteriormente, de ciertos usuarios relevantes de las redes sociales) puede condicionar las decisiones de miles o millones de personas.

En los siguientes apartados de este bloque ampliaremos y estudiaremos todo el marco teórico que nos permitirá esclarecer cuál es este entorno 2.0 en el que los usuarios ejercen como generadores de contenido y, en especial, analizaremos el cómo nosotros sacaremos los datos necesarios que ellos mismos han generado, tomando la red social Twitter como referencia, para hallar los *insights* publicitarios y poder aplicarlos a nuestros intereses y necesidades como publicistas.

1.1.1. Comunicación de masas y medios de comunicación

Consideramos oportuno realizar una apreciación en relación a términos que próximos tienen significados diferentes como es la comunicación de masas (*mass communication*) y los medios de comunicación de masas (*mass media*).

What are mass media? In order to answer this question, we must first back up and define mass communication, which is any form of communication transmitted through a medium (channel) that simultaneously reaches a large number of people. Mass media are the channels that carry mass communication. (Wimmer, 2013:2)^[53].

Siguiendo la definición anterior, entendemos por comunicación de masas toda aquella información que puede llegar desde un único emisor a un número de receptores a gran escala utilizando un medio (canal) concreto. Por ejemplo, una revista en papel puede llegar a tantas personas como ejemplares edite (utilizando como medio la comunicación escrita), ya sean diez o un millón. En el ámbito de la televisión, una cadena puede alcanzar a todos los telespectadores que lo han sintonizado, ya sean de una ciudad, un país o del mundo entero.

En cuanto a los medios de comunicación de masas, nos referimos a las empresas o entidades que, utilizando un medio de comunicación de masas (televisión, radio o prensa) emiten información de carácter noticioso con el potencial de alcanzar a una multiplicidad de receptores que usan el mismo medio. Pensamos, por ejemplo, en un noticiario radiofónico que utiliza como medio las ondas radiofónicas (la radio) para emitir informaciones disponibles para todos aquellos oyentes que sintonicen con dicha emisora.

Como hemos apuntado previamente y seguiremos analizando a continuación, la aparición de nuevos canales de comunicación de masas, y la comunicación a través de Internet en concreto, han ido acompañados de la aparición de nuevos soportes (ordenadores personales, *tablets*, *smartphones*, etc.) y de una serie de posibilidades que han permitido el surgimiento de un nuevo concepto comunicativo que difiere del concepto de medios de comunicación de masas: los medios sociales (*Social Media*).

1.1.2. Los Medios Sociales (*Social Media*)

A diferencia de los medios de comunicación de masas (dónde las llamadas telefónicas del telespectador u oyente o las cartas al director eran las mayores formas de interacción entre usuarios o con el medio de comunicación en concreto), los medios sociales tienen su base en la interacción entre usuarios, más allá de la intervención directa en la temática o el contenido de otros agentes o del mismo servicio que utilizan los usuarios. En resumen, podríamos decir que si un medio de comunicación como *El País* utiliza el medio de comunicación de masas (canal) de Internet para publicar información noticiable a los usuarios que accedan a su web, un medio social, utilizando el mismo canal (Internet), pone a disposición de los usuarios un medio de comunicación social, es decir, enfocado a que sean los propios usuarios los que compartan y difundan la información utilizando su dominio web (o su versión en aplicación móvil), ya sea públicamente o con las personas de su elección.

Si nos remontamos a los inicios de los medios sociales, también definidos como “redes sociales”, se considera a *SixDegrees* como la primera en existir (Ellison, 2007)^[54]. Dicha red social, cuyo nombre remite a la *Teoría de los seis grados* definida en la primera parte de esta tesis, permitía conectar e intercambiar información con otros usuarios, si bien de un modo elemental y básico, con una filosofía que otras redes sociales posteriores tomarían como referente. Le seguirían otros servicios web como los *newsgroups* de USENET y algunos foros (sin ser propiamente redes sociales) como los creados por *CompuServe* (CIS) o los de *American Online* (AOL). Aunque su desarrollo y masificación se harán patentes a mediados de la década de los dos mil, los medios sociales recibieron esta denominación con el notable desarrollo de Internet en el transcurso de la década de los 90. Como características que se derivan del uso que hemos explicado, algunos teóricos consideran que los medios sociales son espacios para la “colaboración masiva” (Tapscott, 2007)^[55], la “cultura participativa” y que dan pie a que pueden regirse por la “inteligencia colectiva”. (Jenkins, 2006)^[56]

Antes de profundizar en el concepto de *Social Media*, es preciso estudiar el concepto de “Web 2.0”, sin el que las redes sociales no serían concebibles.

El concepto Web 2.0 lo acuñó por vez primera Tim O'Reilly, en octubre de 2004, durante la "Web 2.0 Conference"^[57]. Como afirmábamos, será la plataforma básica y necesaria para que las redes sociales, entre otras posibilidades y servicios que permite la Web 2.0, puedan existir tal y como las entendemos a día de hoy. Es importante remarcar que, si bien O'Reilly traza una serie de características y novedades respecto a lo que podríamos denominar "Web 1.0", él mismo afirma que la definición de la "Web 2.0" no puede fijarse ya que está en constante evolución y cambio.

O'Reilly definió, a grandes rasgos, las características de las empresas que han surgido aprovechando las características que ofrece el entorno de la "Web 2.0". Junto a la enumeración, añadiremos algunos comentarios propios explicativos:

- Servicios de software libre, con una escalabilidad rentable

En este sentido, el software libre es aquel que está al alcance de cualquier persona o empresa que decida usarlo y trabajar con él y que cumple una serie de premisas o principios^[58]. En cuanto a la escalabilidad rentable, O'Reilly se refiere a que el coste de incorporar nuevos usuarios al servicio es mucho más bajo que los ingresos que cada usuario genera. Como ejemplo general, podemos afirmar que el coste para Spotify, el servicio de música en *streaming*, varía relativamente poco al pasar de uno a cinco millones de usuarios. Por el contrario, sus ingresos se multiplican por 5.

- El control sobre las fuentes únicas de datos que se hacen más ricos a medida que más personas los usan

De este punto podemos extraer, por ejemplo, que cuantos más usuarios compartan información (del tipo que sea) mayores y mejores datos obtendrá la empresa en cuestión. En referencia a la empresa que nosotros utilizaremos, Twitter, esta tendrá mayor conocimiento del uso que hacen sus usuarios de manera directamente proporcional a la cantidad de información que ellos comparten. Una muestra de ello son los llamados Trending Topic (TT), dato que señala los temas más utilizados del momento por los usuarios de la red social. Solo cuando una masa crítica de usuarios utiliza determinado término este logra situarse como TT^[59]. Este conocimiento, al que empresas y usuarios otorgan un gran valor, reside de forma integral en la empresa, y

puede modificar los criterios a su conveniencia porque solo Twitter controla la fuente (y la fórmula computacional) que determina qué debe ser TT y qué no.

- Considerar a los usuarios como co-desarrolladores

Podemos recurrir de nuevo al caso de Twitter para ejemplificar este punto: antes de la aparición de los hashtags (“#”) tal y como los entendemos hoy, hubo un extenso debate entre los fundadores y algunos usuarios implicados de la comunidad^[60] acerca de cómo podría encontrarse un modo de “agrupar” las conversaciones temáticamente.

- Aprovechar la inteligencia colectiva

Hemos hablado de este concepto anteriormente. Wikipedia es el caso paradigmático de éxito de la inteligencia colectiva. Normalmente, los casos en los que la colaboración entre usuarios aporta soluciones mejores de las que individualmente podrían hallar son más puntuales y concretos.

- Aprovechar la “Larga Cola” (*The Long Tail*⁵) a través del auto-servicio al cliente
- Software por encima del nivel de un único dispositivo

Las tecnologías actuales han hecho que un servicio pueda usarse en una multiplicidad de dispositivos. Del ordenador personal al más reciente *Apple Watch*, todos los dispositivos con conexión a Internet pueden, en mayor o menor grado, permitir el acceso a los servicios que ofrecen las empresas del entorno “Web 2.0”.

- Interfaces de usuario, modelos de desarrollo y modelos de negocio sencillos. (O’Reilly, 2006)^[61]

El caso más evidente lo podemos encontrar en Google y su buscador original. A diferencia de otros competidores, como Yahoo! o Altavista, su simplicidad a la hora de mostrar una interfaz blanca con una línea para ejecutar la búsqueda logró superar la lentitud de las conexiones de aquel entonces. La simplicidad con la que la mayoría

⁵ *The Long Tail*: “Larga Cola” teoría creada por Chris Anderson en octubre de 2004 para ilustrar el cambio producido por Internet y el entorno digital en las leyes de la distribución y las reglas del mercado.

de empresas intentan explicar en qué consiste su servicio es ahora un estándar en la mayoría de los casos.⁶

En las siguientes definiciones seguiremos aportando otras visiones complementarias para enriquecer los conceptos presentados hasta el momento.

Como afirman Prentice y Styhre, el despliegue de posibilidades de la Web 2.0 ha permitido que los medios sociales den lugar nuevas formas de percepción y atención, gracias a los recursos multimedia de los textos, sonidos, imágenes o impresiones táctiles “haptic sense-impression” propias de los medios virtuales. (Prentice, 2005)^[62] (Styhre, 2009)^[63]

Como se deduce de las características esenciales de las redes sociales, el concepto de Web 2.0 (plataforma óptima tanto a nivel ideológico como tecnológico) y el llamado *User Generated Content* (contenido generado por el usuario) son los ejes vertebradores que permiten su aparición, expansión y desarrollo. Asociado *al User Generated Content* encontramos la figura del *prosumer*, descrita en la terminología y que combina sus hábitos de consumo con una comunicación proactiva en sus medios sociales. (Krumm, 2008)^[64]

En la línea de este acercamiento teórico, añadimos la visión de Pepe Tomé, analista de redes y experto en medios sociales, “hasta hace poco estaba claro que los individuos eran receptores y los medios facilitaban el contenido a través de una tecnología, los soportes. Ahora, los cambios tecnológicos permiten al individuo ser generador de contenido, repetirlo, hacerlo suyo y opinar. El contenido ya no es monopolio de unos pocos.” (Tome, 2011:46)^[65]

La aportación siguiente de Miguel del Fresno combina las definiciones que establecen las diferencias entre los medios sociales y los medios de comunicación masiva y hace hincapié en el cambio de relaciones comunicativas entre los individuos y también en su comportamiento como colectivo o ante otras comunidades de individuos: “si los *mass media* son un comunicador profesional, los social media se presentan como un comunicador

⁶ Este es el texto con el que se define Twitter en su página de inicio: *Bienvenido a Twitter. Conéctate con tus amigos y otras personas fascinantes. Obtén actualizaciones instantáneas de las cosas que te interesan. Mira los eventos que se están desarrollando, en tiempo real, desde todos los ángulos.*

interpersonal colectivo. Los social media permiten a los individuos autónomos en sus prácticas comunicativas ser tanto parte del medio como del mensaje en la medida que se les otorga, por parte de otros iguales (*peer-to-peer*⁷), diversos grados de credibilidad a su comunicación individual (uno, a uno) o social (uno, a muchos y muchos a muchos).” (Del Fresno, 2012:101-102)^[66]

Desde el punto de vista técnico, los *Social Media* representan una serie de plataformas y aplicaciones basadas en el acceso a Internet, a partir de las bases conceptuales y tecnológicas de la web 2.0 que facilitan y permiten la creación e intercambio de contenidos generados por los usuarios. (Kaplan, 2010)^[67]

La siguiente definición recoge las principales aportaciones de diferentes investigadores acerca de los Medios Sociales: plataformas, aplicaciones, medios de comunicación online o grupos de aplicaciones basadas en Internet y que son utilizadas por muchos usuarios para su interacción online. Los propios usuarios crean los contenidos y estos contienen todo tipo de contenido multimedia. (Kaplan, 2010)^[67] (López Fernández, 2013)^[68] (Escalona, 2013)^[69]

Existe una confusión acerca del uso del concepto Red Social que podemos resumir para tenerlo presente cuando lleguemos a la definición extendida de Red Social, cuyo concepto desarrollaremos en el apartado 2.3.4.1 Introducción a las aplicaciones de la Teoría de Grafos.

Antes de proseguir, matizaremos un error conceptual de uso común que puede llevar a equívoco. Aunque se usen como sinónimos, “Red Social” difiere del concepto “Plataforma de servicios de las redes sociales”. Twitter, en este sentido, no es una “Red Social”, sino una plataforma de servicios de las redes sociales. Y la aclaración debe ampliarse al uso del término “digital”, porque “Redes Sociales” también pueden producirse en el mundo “no mediatizado” (sin dispositivos electrónicos). Por lo tanto, si queremos ser concretos, cuando hablamos coloquialmente acerca del funcionamiento de una “red social” en Internet, deberíamos tener presente que dicho término es una simplificación demasiado ambigua:

⁷ *Peer-to-peer* (P2P) red de pares, o red entre iguales. En 1999 Shawn Fanning introdujo la aplicación para compartir música y archivos llamada Napster, que significó el comienzo de las redes peer-to-peer.

necesitaremos utilizar el término Red Social Digital. Para la empresa en concreto, utilizaremos “Plataforma de servicios de redes sociales digitales.” (Freire, 2008:585)^[70]

Para obtener mayor claridad, describiremos tanto la Red Social física como la Red Social digital. La “Red Social física” sería aquella red social que se desarrolla sin la ayuda de sistemas electrónicos, como por ejemplo una comunidad de vecinos o un pueblo. Como “Red Social Digital” entendemos aquellas redes que sí operan a través de Internet, ya sea de una manera total o parcial. En su mayoría lo hacen a través de plataformas web que ofrecen los servicios necesarios para que se puedan desarrollar las conexiones entre personas. (Escalona, 2013:55)^[69]

Por otro lado y en contraposición a los medios sociales, tenemos a los medios convencionales de comunicación publicitaria que tradicionalmente se han usado para las inserciones publicitarias en medios de comunicación masivos como la televisión, radio, prensa (diarios, suplemento y revistas), cine, exterior y más recientemente las inserciones publicitarias realizadas en Internet. (García Uceda, 2008)^[71]

Con el fin de resumir y esquematizar las similitudes y diferencias entre los medios de comunicación masivos y los medios de comunicación sociales presentamos la siguiente tabla comparativa:

Medios de comunicación masivos	Medios de comunicación sociales
Participación pasiva del consumidor	Participación activa del consumidor
Comunicación unidireccional de “uno a muchos”	Comunicación bidireccional de “uno a uno”
Contenido elaborado por profesionales de la comunicación informativa	Contenido generado por los usuarios y las marcas (puede incluir la comunicación de

	los medios de comunicación masiva
<i>Message-driven</i> (basado en “mensajes”)	<i>Conversation-driven</i> (basado en “conversaciones”)
La marca tiene el control del mensaje	El control queda repartido entre todos aquellos que emiten, comparten u opinan acerca del mensaje
La comunicación publicitaria trabaja principalmente la imagen de marca	La comunicación publicitaria trabaja principalmente servicios o bienes concretos y la atención y relación consumidor
Métricas: coberturas/frecuencias	Métricas: relación, <i>engagement</i>
Ganchos publicitarios habituales: uso de actores y famosos	Ganchos publicitarios habituales: los propios usuarios implicados y los <i>influencers</i>
Tiempos de impacto comercial: campañas puntuales que desvanecen su notoriedad con el tiempo	Conversaciones y contenido permanece y continúa visible en los buscadores
Creación programada y pre-producida	Creación posible y habitualmente realizada en tiempo real
Estrategia Publicitaria “Arriba abajo” del anunciante al consumidor	Estrategia de “Abajo a arriba” iniciada por el consumidor.
Niveles variables de credibilidad y	Fuertes niveles de credibilidad y

autenticidad	autenticidad en combinación con “ruido” o fuerte presencia de “rumores infundados”
La Inversión publicitaria se realiza principalmente en medios de pago	La inversión publicitaria se realiza principalmente en los medios que utilizan los propios usuarios

Tabla 1: Comparativa de las características de los medios masivos en relación a los medios sociales. Adaptación propia a partir del artículo de la Dra. Angela Hausman. (Hausman, 2012)^[72]

En relación con nuestro ámbito de estudio, vinculado a la elaboración publicitaria, es relevante señalar la creciente tendencia a priorizar cada vez más recursos destinados a la publicidad en medios sociales o a la generación de contenidos publicitarios hechos ex profeso para dichos medios (vídeos virales, por ejemplo). Resultan significativos términos como *Socialnomics*, acuñado por el autor Erik Qualman. El término intenta definir lo que podríamos denominar “la economía subyacente de los medios sociales”, es decir, el flujo de valor entre la inversión en medios sociales y el beneficio que se obtiene de ellos, que alcanza incluso más allá del ámbito económico:

Socialnomics is the value created and shared via social media and its efficient influence on outcomes (economic, political, relational, etc,...). Or, more simply put, it’s word of mouth on digital steroids. A subset of this is that in the future we will no longer search for products and services; rather, they will find us via social media. (Qualman, 2013:XVII)^[42]

El consumo y el modo de acceso a la información clave previa a la compra que realizan los consumidores sigue un proceso totalmente diferente al de la era previa a Internet. Las inversiones en televisión ya no son determinantes para influir de forma decisiva en el proceso de compra de millones de personas. Las recomendaciones de los propios consumidores a otros clientes potenciales han demostrado ser, ya sea a través de las redes sociales o vía las propias páginas que los venden (con esa opción “social” que permite valorar los productos o incluso que los propios usuarios mantengan conversaciones entre ellos acerca de sus características), tan relevantes como la opinión de los expertos más reconocidos^[73]. En este

sentido, lo que dicen, hacen o recomiendan los amigos en Facebook se ha convertido en una de las grandes fuentes de información para la posterior toma de decisiones en futuras compras^[74].

Erik Qualman, además del término *Socialnomics*, utiliza la expresión *Social Media Revolution* debido al cambio radical que han supuesto los Medios Sociales para la comunicación entre personas y empresas. *This revolution is being driven by people and enabled by social media.* (Qualman, 2013:XVII)^[42]

A nivel empresarial, una de las conclusiones a las que nos lleva esta constatación de la preponderancia de los Medios Sociales en la sociedad actual es que la gestión de la presencia de la marca ya no debe ni puede quedar únicamente en manos del departamento de marketing o relaciones públicas. La multiplicidad de mensajes que las marcas deben emitir, la necesidad de la toma rápida de decisiones importantes y la constante actividad de los usuarios en los Medios Sociales hacen necesario que la estrategia de la presencia de marca esté totalmente integrada en la estrategia global de las empresas. Porque, como se ha visto, los Medios Sociales pueden llegar a afectar a todos los departamentos, desde el departamento de atención al cliente a la propia dirección de la empresa.

A modo de ejemplo, apuntaremos varios hechos recientes con notables efectos en las áreas involucradas, desde la política a la economía de una empresa:

La marca automovilística Ford creó “El Movimiento Fiesta”. La acción consistió en dar 100 nuevos Ford Fiesta a modo de prueba a 100 *influencers* en los Medios Sociales durante seis meses. Lo único que hizo la marca fue dejar que los *influencers* hiciesen lo que suelen hacer: tuitear, publicar videos y fotos, y comentar sus experiencias y compartirlas con sus seguidores. Se dejó que los contenidos creados fluyeran a través del sitio web en tiempo real y por Internet, sin filtros ni censura. Los resultados fueron sorprendentes: más de 7.000.000 de vistas de los videos en YouTube, 750.000 visitas de las fotos de Flickr, más de 125.000 me gustas en Facebook y 11.000 reservas de vehículos con una tasa de notoriedad de la campaña superior al 60% y en este caso considerando solo el impacto de esta campaña en Medios Sociales.

Muy recientemente, Pepsi dejó de promocionar su conocido anuncio en la *Super Bowl* de Estados Unidos por primera vez en 20 años. En vez de ello, trasladó su presupuesto millonario a varias acciones de comunicación a través de los medios sociales.

En el ámbito político, tras el archiconocido éxito de Obama gracias a los medios sociales, su cuota de influencia ha avanzado en aumento constante. Como ejemplo más reciente, podemos citar el ascenso de los conservadores de EE.UU. y el llamado *Tea Party* con su *hashtag* de Twitter #TCOT (*Top Conservatives On Twitter*).

En enero de 2009, la primera imagen difundida del mundo del célebre aterrizaje forzoso del avión de *USAirways* en el río Hudson fue a través de Twitter.

1.1.3. Los Medios sociales son Medios de comunicación de masas

En los albores de Internet, parecía que los medios de comunicación tradicionales siempre mantendrían una legitimidad y alcance mayor a los medios de comunicación que surgían propiamente en el entorno digital. Sin embargo, como se ha visto, hay multitud de medios íntegramente digitales, como *El Diario*, *Playground* o *Buzzfeed*, cuyo alcance e impacto es ya notablemente superior al de muchos medios tradicionales equiparables, incluso también al de sus versiones en digital. Los medios sociales, como Facebook o Twitter, han tenido un papel importante a la hora de lograr la consolidación de los medios digitales nativos: el efecto “compartir” permite que una gran cantidad de público se dirija a un artículo en concreto en muy poco tiempo, pudiendo lograr cientos de miles (o millones) de visitas en apenas unas horas. Este tráfico es, en efecto, monetizable, y el coste para llegar a esa cantidad de público puede ser muy bajo si el contenido logra hacerse viral (los usuarios lo comparten voluntariamente y de forma gratuita con sus contactos). (Sánchez Revilla, 2014)^[75] (eMarketer, 2014)^[76]

Otro factor que ha contribuido a esta situación de viabilidad de los medios de comunicación digitales ha sido la concentración de los medios sociales en pocas empresas; de un entorno fragmentado y disperso, con multitud de herramientas, son ahora pocos medios los que manejan la gran parte del tráfico (Facebook, Twitter, Tumblr o Instagram, por ejemplo). De este modo, es más probable que el contenido se difunda con mayor rapidez (sin la necesidad de que el usuario tenga que acceder a otra herramienta ni cambiar de navegador o aplicación). Para la publicidad en los medios sociales, esta situación también ha permitido una mejora considerable: los grandes medios sociales han desarrollado complejas vías para conocer a sus usuarios y poder, así, determinar perfiles muy concretos que ponen luego a disposición de los anunciantes.

Los anunciantes tienen cada vez mayor y mejor acceso al público que les interesa, con una profundidad y precisión más sofisticada que los perfiles que difunden las cadenas de televisión o los periódicos. Recordemos que los métodos de medición de los medios de comunicación convencionales se llevan a cabo a través de encuestas a sus usuarios y su forma

es inevitablemente intrusiva. En cambio, como hemos ido explicando, el usuario de Twitter mostrará, de manera voluntaria, todo aquello que le interesa o le disgusta, ya sea una personalidad pública, un político, un bien cultural o un producto en concreto. Aquí yace el gran potencial de la publicidad en entornos digitales: es posible dirigir los anuncios al nicho de usuarios que nos interesa con una concreción que va desde lo geográfico a los seguidores de una película.

La Dra. Karen Nelson-Field y sus colegas del Instituto Ehrenberg-Bass de Australia planteaban precisamente el siguiente paso en cuanto al uso de la publicidad en los medios sociales: cada vez importa menos el alcance total y el número de seguidores, lo relevante responde a qué tipologías de consumidores alcanza Facebook, cuánto cuesta llegar a cada uno de ellos y qué efectos e impacto queremos causarles.

Facebook is becoming more and more like traditional media. It may be time for advertisers to move on from worrying about how many fans they have to instead explore how many category buyers Facebook can reach, for what cost, and to what effect. (Nelson-Field, 2012:2)^[77]

Medir la audiencia en términos estrictamente cuantitativos es menos eficiente que analizar cómo es esa audiencia y a qué estímulos responde. En este sentido, de poco nos sirve saber que nuestra marca es *Trending Topic* en Twitter y que uno o tres millones de personas están hablando de ella, lo que realmente nos aportará valor y nos permitirá aprovechar el contenido de lo que se habla (para participar en la conversación, obtener información para el futuro o modificar errores cometidos) es *entender* qué es lo que se dice. Uno de los beneficios que podríamos conseguir con ello es llegar a dilucidar los *insights* que nos permitirán, en un futuro, seguir llegando a nuestros clientes con el mensaje adecuado y adaptándonos o adelantándonos a sus necesidades. Escuchar las conversaciones puede llevarnos, también, a aquellos consumidores potenciales que se muestran indecisos o muestran reservas. Los *insights* pueden permitirnos hacer una publicidad concreta destinada exclusivamente a alcanzarles por primera vez de forma que sea importante para sus intereses.

2.- LA INVESTIGACIÓN EN LA COMUNICACIÓN PUBLICITARIA

2. La investigación en la comunicación publicitaria

2.1. El *insight* en la comunicación publicitaria

La definición de *insight* supone un reto, ya no solo para lograr una traducción precisa al castellano, sino también para concretar su significado original en inglés. Ello se debe, principalmente, a su connotación “etérea”, queriendo decir con ello que un *insight* no es medible ni responde a ninguna entidad física o plausible en palabras exactas.

No obstante, en vez de limitarnos a su comprensión, hemos realizado una extensa búsqueda de las definiciones más relevantes y completas que se han proporcionado del *insight* a lo largo del tiempo, aportando también definiciones que se complementan entre sí. Aspiramos así a dar una visión global, rica y poliédrica de todas sus implicaciones, desde un nivel conceptual hasta su aplicación en el campo del marketing y la publicidad. Cerraremos el apartado de las definiciones con varios ejemplos prácticos llevados a cabo por empresas importantes. Si bien su significado es complejo, su valor en términos publicitarios es determinante, puesto que, como veremos, de su obtención radica la posibilidad de lograr realizar comunicaciones publicitarias con un alto potencial de éxito entre los consumidores. Por eso, los métodos de obtención del *insight* son sofisticados y extensos, y a ellos les dedicaremos un notable apartado que cierra este punto del capítulo. Como hemos ido desarrollando, con la aparición de los Medios Sociales y, en concreto, con el desarrollo de las redes sociales, las posibilidades de lograr nuevas vías para encontrar *insights* han crecido exponencialmente y, precisamente, este es el objetivo principal que nos ocupará en los siguientes capítulos de esta tesis.

El término *insight* fue estudiado por primera vez por la *Gestalt Psychology* a inicios del siglo XX (Mayer et al., 1996)^[78]. En el campo de la psicología *insight* recibiría posteriormente constantes aportaciones a las definiciones originarias. En cuanto a nuestro estudio, el acercamiento será principalmente el que se ha dado al término en los campos del marketing y la publicidad. Confiamos alcanzar así la más amplia y rica concepción del término que nos permita avanzar en su método de obtención y estudio.

Insight carece de traducción literal o directa al castellano, salvo la que encontramos en el diccionario Akal de psicología, en el que *insight* es definido como “compreensión súbita”. La expresión aspira a definir satisfactoriamente el término inglés *insight* junto al alemán *einsicht*, aplicado por vez primera por W. Köhler en 1917. Sin embargo, aclararemos que lo súbito no es una característica esencial del entendimiento, sino que se considera como una reanudación de la actividad observable a la hora de la ejecución de una solución (Doron, 1998:112)^[79]. De todos modos, si bien podemos hallar “compreensión” en la terminología del ámbito de la psicología, el término no se encuentra definido en la Real Academia de la Lengua Española. Por ello, en nuestro estudio, preferimos utilizar, por su apropiado acercamiento al significado empleado en el marketing y la publicidad del término *insight*, el término “entendimiento”.

En cuanto al la condición de hallazgo “súbito”, en contraposición al hallazgo fruto de una investigación racional, extensa y siguiendo un procedimiento científico deductivo, el autor, desde el campo de la psicología, identifica el término *insight* como una “súbita reorganización de la percepción de los elementos que configuran el problema, de modo que la nueva relación entre los elementos conduce a la solución”. (Valdivieso, 2002)^[80]

Si avanzamos en la definición de *insight*, es interesante incorporar la definición de la *Webster's Encyclopedic*:

An instance of apprehending the true of a thing. 1. Though intuitive understanding: an insight into the life of the village. 2. Penetrating mental vision or discernment; faculty of seeing into inner character of underlying truth: a man of great insight. (Publishing, 1996:735)^[81]

De esta definición señalamos la presencia en un *insight* del hallazgo de algo “verdadero”, algo que alcanza la raíz o la parte esencial de lo que se está intentando encontrar.

Situado el *insight* como un entendimiento súbito de algo relevante, alejado de los aspectos superficiales de un descubrimiento y asociándolo a algo de “mayor profundidad”, procedemos ahora a adentrarnos en sus aplicaciones en el campo de la publicidad y el marketing.

Por ejemplo, la definición que ofrece Bullmore, presidente de *The Market Research Society*, nos ayuda a comprender la concepción del *insight* y de su utilidad en el campo de la estrategia de marca:

A new understanding probably of human behaviour or attitude, as a result of which action must be taken and an enterprise more efficiently conducted. (Bullmore, 2006:141)^[82]

Destacamos de la definición de Bullmore la referencia al descubrimiento de alguna “actitud, posicionamiento o comportamiento” de las personas. En este sentido, entendemos que este descubrimiento podremos intentarlo aplicar a la solución que puede ofrecer el producto de nuestra empresa.

Bill Bernbach⁸, destacado creativo publicitario, ofrece por su lado ese acercamiento que explicábamos acerca de la “profundidad” del descubrimiento, en contraposición a la superficialidad o poca relevancia del hallazgo:

At the heart of an effective creative philosophy is the belief that nothing is so powerful as an insight into human nature, what compulsions drive a man, what instincts dominate his action, even though his language so often camouflages what really motivates him. For if you know these things about [a] man you can touch him at the core of his being” Obtenido del discurso realizado por Bill Bernbach en 1980 ante la American Association of Advertising Agencies. (Steel, 1998:XIII)^[83]

⁸ William Bernbach (1911 - 1982) publicista estadounidense, fundador de la actual DDB Worldwide.

Siguiendo esta línea de razonamiento, Johannes Hartmann, vicepresidente de *Consumer & Market Insight Foods* de la empresa Unilever en los Países Bajos, durante la conferencia “*Consumer Insight*” celebrada en Milán en 2007, hizo referencia al *insight* desde la perspectiva de la verdadera naturaleza humana:

Another response to the question “what is insight” is more acceptable, more tractable, and brings us closer to the humanistic aspect of insights which this paper proposes. The notion is that the idea of insight has to be seen as fundamental. Insights, human truth, big ideas are fundamental business beliefs that are “large” in scope and at the same time have the potential to make the world a little better. (Hartmann, 2007)^[84]

Hartmann aporta en su definición un aspecto que también debe ser tenido en cuenta en los ámbitos de la publicidad y el marketing: al decir que un *insight* es algo fundamental, su utilidad aplicada en el campo publicitario debe ser útil y no gratuita o trivial. En efecto, será el deber de los publicistas ofrecer una respuesta práctica, positiva y valiosa ante el *insight* descubierto.

Nos permitiremos aquí un breve apunte a una conocida expresión popular que una persona expresa al hallar una clave que resuelve un problema o situación que antes era incapaz de afrontar: “¿Cómo no me había dado cuenta antes?”. Esta pregunta, de uso familiar y cotidiano, la mencionamos en relación a la siguiente definición de Laybourne. En el significado popular, esta pregunta suele asociarse a la estupefacción que uno siente por haber tenido la respuesta muy cerca desde que se planteó el problema pero, paradójicamente, era incapaz de encontrarla hasta que de golpe “cae en la cuenta” de su solución.

La definición de Laybourne dice así: “insight is described as an observation, or piece of information that hasn’t been noticed, but has to be fresh and inspiring and only obvious in retrospect”. (Laybourne, 2007:33)^[85]

En efecto, el *insight* es inspirador e inédito y, como decíamos, solo nos parecerá obvio una vez lo hayamos encontrado. Al expresar “¿Cómo no me había dado cuenta antes?” queremos decir que, en efecto, la solución era obvia, pero hasta ese momento nos parecía inaccesible.

A continuació citamos la respuesta del TermCat ante una consulta sobre la definició de *insight*:

El terme anglès *insight* pertany principalment a l'àmbit de la psicologia i de la filosofia i es tradueix sovint per intuïció o introspecció segons el context. L'expressió consumer *insight*, potser derivada d'aquest sentit originari, és, però, un concepte nou que es fa servir en l'àmbit de la publicitat per a referir-se a la comprensió o la captació intuïtiva del pensament del consumidor (les seves idees, els seus comportaments, etc, allò més profund del pensament que potser ni ell mateix sap o reconeix). En les obres lexicogràfiques i terminològiques de l'àmbit de la publicitat no hem documentat aquest terme traduït a cap llengua (ni en francès ni en espanyol). És per això que només podem oferir-vos una forma descriptiva per a traduir-lo. La nostra proposta seria "coneixement profund (del pensament) del consumidor", que és la forma que s'utilitza a vegades en la paràfrasi d'aquest concepte. (Balanzó⁹, 2011:61)^[86]

Al referirnos a esa sensación de estupefacción ante el hallazgo de un *insight*, produciéndose un alivio de la presión que el problema o la cuestión a resolver nos causaba, encontramos una carga emocional que podemos asociar al sentimiento de liberación y satisfacción. De Nicolás lo describe de la siguiente manera:

La vivencia del *insight* es importante a la hora de clarificar este concepto. Aquellas personas que en algún momento han producido *insight*, lo definen como llamaradas fugaces de la inteligencia, iluminaciones anticipadoras que validan una solución determinada tras una rápida revisión crítica. (Nicolás Carrillo, 1999:146)^[87].

Johannes Hartmann, al que hemos citado anteriormente, realiza en la siguiente cita ese nexo que necesitamos entre el descubrimiento "profundo" del *insight* y nuestro producto que como publicistas queremos ofrecer como solución.

In many companies the operating point of view is that real need of insight is for the explicit positioning of the product or service, a utilitarian value of insight. The insight as a key to understanding basic "human truth" is ignored, in the service of business expediency. (Hartmann, 2007:114)^[84]

⁹ Cristina Bofill. Servei de Consultes del Termcat.cat. Consulta realizada por correo electrònic a petició de Cristina de Balanzó el 24/07/2006 a las 12h42'.

De la intersección entre *insight*, emociones y publicidad, la investigadora López Vázquez señala acertadamente su posible relación:

Para hablar de *insights* tenemos que pensar en necesidades, expectativas, frustraciones, que la publicidad canaliza en productos y marcas para generar satisfacciones racionales y emocionales a nivel consciente e inconsciente, donde las firmas se convierten en ese mago que es capaz de convertir los sueños en realidad. (López Vázquez, 2007:40)^[88]

López Vázquez, aplicando íntegramente el *insight* al terreno del marketing y la publicidad, lo sintetiza de la siguiente manera:

“Los *insights* son, en definitiva, verdades sobre el consumidor...” (López Vázquez, 2007:42)^[88]. Mike Lee intenta resumir en una expresión que un supuesto consumidor exclamaría al ver un *insight* resuelto satisfactoriamente por una campaña de publicidad asociada a un producto determinado: “*This brand understands me! That is exactly how I feel!*” (Lee et al., 2008:171)^[89]. En efecto, un *insight* bien aprovechado puede tener el poder de cambiar el comportamiento del consumidor y de este modo beneficiar a la marca y alcanzar los objetivos que persigue el marketing. (Beckley et al., 2012)^[90]

Otro aspecto fundamental reside en la importancia que tiene el cómo se expresa el *insight*, ya que de ello dependerá su potencial para inspirar.

Product satisfaction arises less from inherent construction and performance than from consumers internalized perceptions of personal utility” in a low-potency insight.

People don’t want quarter-inch drills. They want quarter-inch holes” is a high-potency insight. (Bullmore, 2006:141-142)^[82]

Anteriormente señalábamos en la definición de Hartmann la importancia del compromiso que debe tener el publicista o el experto en marketing en cuanto a la utilidad y el valor real de su propuesta ante el hallazgo de un *insight*. Un *insight* no es meramente curioso o interesante, debe aportar información y una acción concreta que lo resuelva:

While findings are free to be trivial and merely interesting (or not) insights bear a much greater responsibility. Insights must be informative and actionable, and straightforward enough that our non-research peers can easily digest them. (Williams, 2007:10)^[91]

Para complementar el tejido de definiciones que hemos presentado por el momento, nos referimos ahora a las marcas y a sus concisas y lúcidas definiciones que ofrecen:

Diageo:

“A penetrating discovery about consumer motivations, applied to unlock growth”

General Mills:

“An insight is something you know that your competitors do not”

Cadbury Schweppes:

“Getting to the heart of why people do what they do, and using that knowledge to help us grow”

Reckitt Benckiser:

“New, creative opportunity derived from a profound understanding of the consumer. It describes how we might add new value to consumers through brands”

Unilever:

“The element all of you know about the consumer on which the brand is founded....insights go deeper and have an interpretative twist that helps connect the brand to the consumer”

Coca Cola &Co. :

“An interpretation of an observation that has relevance to my business, anyone can observe, but only the good can interpret, and is so doing create huge relevance and value” (Laybourne¹⁰, 2007:33)^[85]

¹⁰ Citado por Cristina de Balanzó en su tesis doctoral. Definiciones recogidas por el investigador Laybourne y expuestas en la conferencia de apertura del Congreso Internacional sobre *Consumer Insight* organizado por ESOMAR. Laybourne P. “The use and abuse of insight...a question of truth”. [CD-ROM]. ESOMAR WORLD RESEARCH CONFERENCE. *Consumer insight 2007: Milan 7-9 May*.

Hemos visto, a lo largo de esta serie de definiciones y apuntes acerca de la definición, los usos y la finalidad de los *insights*, cómo el término en sí mismo y su vinculación al proceso creativo del publicista despiertan un gran interés y curiosidad por parte de todos aquellos que se dedican a su estudio. Ello se debe, según nuestra consideración, a que para su elaboración se mezcla el análisis de la información junto con cierto momento de “inspiración”. Cerrábamos las definiciones con la que aporta Coca Cola, y que dice que es “ver aquello que todo el mundo puede ver pero que solo unos pocos entendidos pueden interpretar”. En efecto, el término “entendidos” nos resume esta admiración por el proceso de exploración de los *insights*, supuestamente solo al alcance de aquellos que saben combinar la información necesaria con un aspecto creativo asociado a la intuición.

2.1.1. El *insight* creativo

Como quizás ha quedado implícito a lo largo de las anteriores definiciones y apuntes que hemos presentado, un aspecto asociado al *insight* es la trascendencia del proceso creativo que lo permite. En este sentido, aportaremos algunas consideraciones teóricas:

Dentro del proceso creativo y eminentemente en el campo filosófico, el *insight* se asocia al término griego *Eureka* o a la interjección coloquial: “Ahá!” (Gardner, 2006)^[92] o al término “Serendipia”. En otras palabras, se define como “*descubrimiento fortuito: encontrar cosas cuando no se están buscando*” (Romo, 2009:55)^[93]

Otros autores también han analizado el *insight* desde el proceso creativo, como el libro *Perspectives in Creativity* y su mención a Sydney J. Parnes: “Suppose we started with the idea that the essence of the concept of creativity is the fundamental notion of the “aha” - meaning the fresh and relevant association of thoughts, facts, and ideas, into a new configuration which pleases, which has meaning beyond the sum of the parts, which provides a synergic effect”. (Taylor, 2009:225)^[94]

Csikszentmihályi lo aborda desde una perspectiva similar y complementaria. Como vemos, en relación a su condición “etérea” que apuntábamos al inicio de este apartado, el uso de analogías y metáforas es especialmente útil para visualizar la relevancia del descubrimiento intangible:

La intuición tiene lugar, presumiblemente, cuando una conexión inconsciente entre ideas encaja tan bien que se ve forzada a salir de la conciencia, lo mismo que un corcho mantenido en el agua sale y salta en el aire cuando se suelta. (Csikszentmihalyi¹¹, 1998:131)^[95]

¹¹ Csikszentmihalyi, Mihaly. (1934) Profesor de la universidad de California, ha destacado por sus estudios en Psicología Positiva y las Neurociencias.

2.1.2. El *consumer insight*

Retomamos ahora la definición que TermCat ofreció a Cristina de Balanzó, concretamente la mención al término *Consumer insight*. La finalidad de esta definición es poder completar de una manera global y con todos los matices imprescindibles la visión del *insight*, cuyas nuevas vías de obtención son el objeto principal de esta tesis.

Como su propio término indica, *consumer insight* sitúa al consumidor como principal destinatario del *insight*. Ello otorga a la publicidad y el marketing el protagonismo de realizarlo adecuadamente.

Sería en el año 1992 que Lisa Fortini-Campbell, en su texto *Hitting the Sweet Spot*, introduciría el término *Consumer insights* en el vocabulario del marketing por primera vez. Se iniciaba así una creciente sensibilización por parte de las agencias y los creativos publicitarios acerca de la relevancia que podía tener una investigación perspicaz, estructurada y profundamente analizada a la hora de producir publicidad creativa y eficaz en términos de ventas (Fortini-Campbell, 2001)^[96]. Sin el conocimiento del consumidor y aquellas necesidades o inquietudes que todavía no ha podido resolver (incluyendo especialmente aquellas de las que quizás todavía no es consciente), esta tarea será inalcanzable. En efecto, el *consumer insight* es el resultado del trabajo riguroso e investigativo de las agencias y los publicistas.

Esquemáticamente y como tentativa de definición y consideraciones para trabajar con los *consumer insight*, destacamos las doce características, consejos o apuntes que Fortini aporta acerca su elaboración:

1. “*Insight* del consumidor + *insight* de la marca = *sweet spot*”. De la suma del *insight* del consumidor y el *insight* de la marca puede salir un spot “dulce”, entendiendo por ello su agradable y buen impacto en la audiencia.
2. Con el hallazgo de un *insight*, será el propio consumidor el que quedará persuadido por sí mismo:

- “*Breakthrough happens not when the advertiser breaks out the multitude of advertising messages, but when the consumer breaks into the message because he or she sees something meaningful or relevant*” (Fortini-Campbell, 2001)^[96]
3. El primer paso para hallar un *consumer insight* es ponerse en la piel del consumidor.
 4. Para entender a los consumidores, uno debe primero entender también sus propias necesidades y aspiraciones.
 5. Los *insights* deben responder siempre a hechos concretos.
 6. Se debe Considerar esta cadena de valor dentro de la obtención de un *insight*: “*Data-information-insight-inspiration*”
 7. Las pequeñas pistas pueden revelar grandes *insights*:

The little clues people leave about themselves can show you to connect with them. As we mentioned first section, the more you know about your consumers, the better you can make your brand intersect their paths, rather than making them walk over to yours. (Fortini-Campbell, 2001)^[96]
 8. Se conocerá mejor a los consumidores en su entorno cotidiano.
 9. Es importante Interactuar con los consumidores, elaborar preguntas detenidamente, escuchar lo que dicen y tenerlo en cuenta.
 10. Los *insights* deben ser expresados con claridad y sencillez.
 11. Es imprescindible realizar hipótesis acerca de los hallazgos y estos deben confirmar los hechos a los que se refieren.
 12. Hay que actuar en defensa y los intereses del consumidor. (Fortini-Campbell, 2001:163)^[96]

A través de las definiciones de *insight* y del término *consumer insight* entendemos con mayor claridad el qué y el cómo debemos enfocar el proceso de su investigación y obtención. Phillip Kotler se refería a ambos términos de la siguiente manera: “*We need to change into a more consumer based (centered) perspective. The magic word is consumer insight*”.

En el ámbito académico y de estudio de la publicidad hallamos más comentarios acerca de la ascendente necesidad y pertinencia de privilegiar la búsqueda de los *consumer insight*:

Consumer insight es una pieza clave y fundamental en la nueva era del marketing, ya que será el factor que evalúe el éxito, tanto de las marcas como de las compañías, además de ser el verdadero valor y el diferencial de la investigación de mercados, en sí mismo. Como dice Dona Vitale en su libro sobre Consumer Insights 2.0.: "Since 1992, ... Today, we need consumer insights more than ever before. (Vitale, 2006:3)^[97]

Como se ha podido constatar, el *insight* se ha analizado desde disciplinas diferentes y con diferentes definiciones, que todas ellas, reflejan la dificultad de obtener un consenso dentro del ámbito del marketing y la publicidad. El *insight* es un concepto que va tanto del lado de la intuición y la imaginación como de su necesidad, cada vez más creciente, de una fuente fiable de información, de los datos de la investigación y de la aplicación de un proceso de análisis realizado de forma racional y estructurada.

Podemos afirmar que el proceso de elaboración de *insights* requiere de varias fases todas ellas imprescindibles y que empiezan por la obtención de información que se ha ido acumulando durante un proceso de observación, continua por una fase de análisis para alcanzar un conocimiento nuevo o diferente de aquello que está estudiando. Fruto de la síntesis y análisis estableciendo relaciones entre el consumidor y la marca, aporta una claridad y una comprensión que permite la identificación del *insight*.

Algunos lo consideran un descubrimiento, el *aha!* o *eureka!*

El término *insight* alberga diferentes conceptos en él mismo: ideas, motivaciones, vivencias, sentimientos, necesidades, frenos,.. A partir de los cuales se puede construir un conocimiento nuevo, una emoción que sirva de desencadenante de una idea, o de una acción de comunicación concreta, y que actúe de guía para el posicionamiento de una marca. Ha de servir para alcanzar un fin o un objetivo. (Balanzó, 2011:51)^[86]

2.1.3. Aclaraciones acerca de los *insights*

Pese a ser un concepto que hemos calificado de “etéreo” y que, por su propia definición se nos presenta como abstracto, ofreceremos aquí algunas aclaraciones que nos ayudarán a dilucidar qué es un *insight* partiendo, precisamente, de los equívocos a los que puede llevar un mal entendimiento del término.

Cristina de Balanzó recogió en su tesis doctoral un estudio realizado por la empresa de bienes de gran consumo Johnson & Johnson. La empresa elaboró un programa de desarrollo de *insights* al que llamó *Marketing insights training* y fue ofrecido por su Marketing Manager de la división de *Consumer Healthcare España*, Pablo Galiana. Galiana destacó algunos aspectos que frecuentemente pueden ser susceptibles de ser confundidos con *insights* y que mostró de la siguiente manera. A continuación analizamos algunos ejemplos:

Un *insight* no es una creencia aceptada por el consumidor: “Como madre quiero que mis hijos tengan una buena nutrición, así tendrán mayores probabilidades de estar sanos y de que les vaya bien en el colegio”.

Si nos remitimos a las definiciones ofrecidas de *insight* veremos que aquello de lo que el consumidor ya tiene conciencia, sea una necesidad o un deseo, no pertenece a la categoría de *insight*.

Ni una idea de producto: “Un producto que me dé la protección de una compresa, pero la discreción de un salvaslip”.

Lo importante de un *insight* es que remite al hecho y no a la idea. Es decir, lograr ofrecer un producto determinado de una manera adecuada sí puede ser un *insight*, pero no será suficiente alcanzar una idea elaborada.

Tampoco es un dato indiscutible: “El 52% de las mujeres que usan tampones y compresas durante la regla usan las compresas por la noche”.

Este dato estadístico, si bien puede ser tenido en cuenta como información para investigar acerca de productos o necesidades concretas, no es un *insight*. Un dato estadístico es, en todo caso, el resultado de la elaboración de un estudio y, por lo tanto, aunque no sepamos con certeza su valor, sí hemos previsto las respuestas posibles (“las utilizo de día”, “de noche”, “indistintamente”) y, por lo tanto, nos darán información relevante pero no nos llevarán por sí mismos a una solución nueva.

Un *insight* no es una afirmación de una necesidad: “La piel del bebé necesita un cuidado especial frente a los efectos de secado de la piel después de un baño”.

Si bien esta afirmación es un hecho, su mera constatación no es un *insight*. Caemos de nuevo en una constatación evidente y que requiere solo requiere una mera observación (piel dañada) que tiene una solución ya existente (aplicar una loción adecuada).

Tampoco es un deseo que expresa la propia empresa: “Tener que hidratarse todos los días lleva mucho tiempo y esfuerzo y uno pasa frío en el proceso. ¿No sería genial si pudiéramos hidratarnos en la ducha?”

La afirmación anterior viene acompañada de una pregunta que se plantea como solución posible (apelando a un nuevo producto o jabón, etc.) y, por lo tanto, deja de ser un *insight* si su búsqueda ya está delimitada.

Ni un deseo del consumidor: “Me encantaría que mi crema de manos hiciera que mis manos se vieran más hidratadas”

Este caso es parecido al que abría estos comentarios y que mostraba la creencia de un consumidor. Aquí sucede algo parecido: si el cliente expresa explícitamente su deseo, nuestro trabajo será ofrecerle una forma de satisfacerlo pero ello no convierte su afirmación en un *insight*. (Balanzó, 2011:69)^[86]

Con el estudio de estos ejemplos podemos concluir que los *insights* pueden explicar por qué la gente actúa de una determinada manera pero que ello no significa que se dediquen a describir comportamientos ni a suscribir afirmaciones o necesidades; su cometido es ayudar a encontrar esos comportamientos y así dar pie a las posibles problemáticas, necesidades o deseos que se desprenden de su estudio y que el publicista aspira a resolver. Aquí radica la

idea central del *insight* que nos ha acompañado desde el inicio y que consolidamos en este cierre: un *insight* puede llegar a provocar nuevas formas de generación de demanda de un determinado producto o servicio. El “súbito entendimiento” que nos proporciona un *insight* es lo que nos conducirá a realizar una determinada acción y en su éxito constataremos el cambio de comportamiento de los consumidores gracias a que han apreciado y comprobado la utilidad de nuestra propuesta.

2.1.3.1. Ejemplos de *insights*¹²

Mostraremos aquí una serie de anuncios muy conocidos y añadiremos un comentario para desarrollar el significado.

Nike Air Max 360

Spot: Los grandes sufren

Observación: El *cushing* (síndrome de Hiperadrenocorticismos) reduce el riesgo de dolor en la práctica del deporte.

Insight: Puedes tomar el camino del dolor físico, pero nunca puedes evitar el dolor emocional del deporte (dar poder al consumidor).

Comentario adicional: las zapatillas deportivas Nike Air Max se ocupan de protegerte de todo el dolor físico que el deporte pueda causarte y dejarte el camino libre para que tu victoria sea también el fin del dolor emocional.

Nike Women

Spot: Baile en el estadio

Observación: Bailar no está considerado como un deporte o como una aspiración atlética válida.

Insight: Las bailarinas (y las mujeres en general) están indignadas al ser subestimadas por hombres ignorantes.

¹² Los casos comentados pertenecen a las agencias Weiden+Kennedy (Nike, Honda, Coca Cola), McCann Erickson (V&T) y SCPF (BMW y IKEA) y a trabajos realizados en 2006 y 2007

Comentario adicional: La línea de ropa y accesorios *Nike Women* entiende y, más allá de eso, reconoce la importancia capital de las mujeres en el deporte. Con sus productos apoyará todo aquello que ellas consideren deporte sin que una visión patriarcal lo limite a un determinado número de “deportes reconocidos”.

Coca Cola

Spot: *The coke site of life*

Observación: Solemos mirar el lado negativo de las cosas con mayor frecuencia que el positivo (en el lenguaje humano existen 7 veces más palabras con connotaciones negativas que positivas).

Insight: La felicidad está en tus manos, tómala.

Comentario adicional: una de las frustraciones habituales que tienen las personas es la sensación de que alcanzar sus sueños no está, precisamente, al alcance de su mano. Por contra, el placer de tomar Coca-Cola sí nos proporciona unos instantes de felicidad y está, literalmente, en nuestras manos.

IKEA

Spot: Bienvenido a la república independiente de tu casa.

Observación: Sólo en casa las personas hacen realmente lo que quieren y con las normas que ellos mismos dictan. Esta libertad, en cambio, queda modificada por otros códigos sociales cuando uno sale de su casa.

Insight: Es en la intimidad de tu propia casa cuando más libre te sientes.

Comentario adicional: Habiendo descubierto la importancia que tienen algunas libertades que uno se permite en su propia casa y que no puede realizar en ningún otro lugar, la labor de IKEA es proporcionar los mejores muebles y accesorios para que el ejercicio de esa libertad pueda ser puesto en práctica en las mejores condiciones.

BMW

Spot: *Be water my friend*

Observación: las personas que compran un 4x4 no suelen usarlo en el campo sino principalmente en zonas urbanas.

Insight: No hay nada tan simple y adaptable como el agua.

V&T de Nestea

SPOT: Nuestro secreto

Observación: lo que gusta a las mujeres es la belleza natural y, por ello, el cuidado personal lo consideran un secreto.

Insight: las mujeres no admitirían que se cuidan ni en un vuelo de la CIA. La belleza natural es lo que más atrae.

Comentario adicional: por sus cualidades naturales, las mujeres que toman V&T de Nestea no hace falta que admitan que se están cuidando. Por el mero hecho de tomarlo, se están cuidando, sin necesidad de explicitarlo y sin el temor a ser “descubiertas”. (Balanzó, 2011:106-114)^[86]

A modo de resumen, desde sus orígenes en el terreno de la filosofía y lo que hemos convenido en definir como “súbitos entendimientos”, hemos visto el complejo proceso de maduración del *insight* y valorado todos aquellos matices que se han desarrollado entorno a su definición. Este trabajo ha sido enfocado siempre con el destacado lugar actual que el *insight* ocupa en nuestro campo de estudio: la publicidad y el marketing. Como hemos visto a través del *consumer insight*, situar al consumidor en el epicentro de las investigaciones ha sido prioritario, puesto que nadie mejor que él conoce sus problemas, inquietudes y necesidades. No obstante, quizás nadie mejor que los publicistas saben cómo hacer inteligibles y conscientes sus aspiraciones, ofreciéndole al consumidor, con información y hechos, la solución o el *eureka* que este sabrá apreciar si la investigación y el trabajo se han llevado a cabo con profesionalidad y rigor. Para ello, como apuntaremos en los próximos apartados, recabar información a través de distintos métodos de investigación es una parte necesaria del proceso. Aunque por definición se encuentre asociado a la espontaneidad o a la superación de un razonamiento lógico por medio de la intuición, el caldo de cultivo de un *insight* es la observación atenta y la obtención de datos de calidad que puedan ser puestos en relación.

2.2. Métodos de investigación aplicados a la comunicación

En esta sección realizaremos un estudio y categorización de los principales métodos de investigación utilizados para el desarrollo de campañas de comunicación. Como punto de partida señalaremos que, en función de la fase de la campaña, se utilizan distintos métodos y con diferentes finalidades. A grandes rasgos, los métodos de investigación que se aplican durante la creación de una campaña de comunicación se utilizan o bien en el proceso de obtención de documentación e información previos al inicio de cualquier decisión efectiva, o bien una vez se ha producido la campaña pero se quieren medir su eficiencia y efecto antes de emitirla públicamente. En el primer caso, los datos obtenidos de las investigaciones permitirán conocer con mejor detalle ciertos aspectos que incidirán en la toma de decisiones de la campaña. En el segundo caso, el objetivo perseguido es asegurar, en la medida de lo posible, que la campaña que se ha realizado sea recibida positivamente en su emisión gracias al estudio de la reacción y las valoraciones que un determinado grupo de personas nos proporcionarán en exclusiva.

Por otro lado, encontramos también distintos métodos de investigación que se utilizan con posterioridad o en el transcurso de la emisión de la campaña. Su objetivo persigue, en estos casos, la medición del éxito de la campaña y del impacto alcanzado con su lanzamiento.

Por la propia naturaleza de nuestra investigación, nos centraremos en los métodos y técnicas que los anunciantes y agencias utilizan antes de la creación de las campañas, en concreto antes de empezar la labor propiamente creativa. Nos ubicaremos en la situación anterior al inicio de ese trabajo creativo, ya sea antes de la labor que realiza el anunciante en la conceptualización del producto o servicio o antes de que agencias y publicistas se dispongan a trabajar para conceptualizar y crear la campaña de publicidad. Efectivamente, será antes de empezar con la labor creativa que los métodos de obtención de *insights* jugarán un papel primordial, pudiendo aportar las claves que nos revelen qué dirección seguir en su desarrollo posterior.

Dentro de las técnicas de investigación que hemos estudiado hallaremos dos categorías: las técnicas cuantitativas y las técnicas cualitativas. Además de señalar las más relevantes de ambas categorías, incorporaremos en esta sección algunas metodologías específicas que han desarrollado varios anunciantes y agencias reales y que aplican con el fin de obtener *insights*.

2.2.1. Métodos tradicionales de investigación

Llamamos métodos tradicionales de investigación a aquellos cuya práctica lleva siendo regular desde antes de la aparición de Internet y que están basados, principalmente, en herramientas de medición analógicas, aunque se aprovechen las nuevas tecnologías en su aplicación. Detallaremos, en los diferentes casos, las ventajas y los inconvenientes que plantea su uso.

Los principales métodos cuantitativos son las encuestas, los cuestionarios y los llamados “análisis de contenido”.

En lo que a métodos cualitativos se refiere, situaremos como destacados las entrevistas (en profundidad, telefónicas, por correo o por Internet), los *Focus Groups* y el método Delphi.

2.2.1.1. Métodos cuantitativos

La perspectiva de los métodos cuantitativos engloba las técnicas de investigación que buscan obtener y medir datos sobre una realidad social concreta. Son métodos empíricos, basados en los datos que se obtienen a través de una investigación (datos primarios) o bien a través del análisis de datos preexistentes (datos secundarios). Los métodos de análisis de datos preexistentes son más habituales en las encuestas y los sondeos y su fin es alcanzar conclusiones y datos cruzados que aporten un conocimiento más completo de la investigación. (Berganza Conde, 2005)^[98]

Se podrían definir las técnicas cuantitativas como aquellas que buscan resumir algunos hechos de la realidad numéricamente. Proviene del paradigma positivista y son recomendables cuando el caso requiera un análisis en extensión y generalizador, es decir, que cuyos resultados sean en cierto modo “escalables” a partir de una muestra que cumpla las condiciones de aleatoriedad (si es el caso). Por los resultados que ofrecen, enfatizan los análisis causales y una validación concreta de la teoría. Como apuntamos, las técnicas más utilizadas son la encuesta y las fuentes de datos estadísticos. (Manheim, 1988:15)^[99] (Alvira Martín, 1991:54)^[100]

2.2.1.1.1. La encuesta

Las encuestas se utilizan para la recogida de información y constituyen una de las técnicas de investigación más conocidas y empleadas. Ampliamente conocidas también entre el público, una encuesta plantea una serie de cuestiones y recoge las respuestas de cada uno de los encuestados. (Berganza Conde, 2005)^[98]

Uno de sus usos más reconocidos es el que sitúa a las encuestas como el procedimiento estadístico que permite captar la opinión de una sociedad o de un grupo social respecto a un tema en concreto y, con ello, es posible dilucidar el sentido y la intensidad de las corrientes de opinión mayoritarias. Las encuestas, en este sentido, pueden captar tanto situaciones y hechos como opiniones. (Monzón Arribas, 1987:164)^[101]

Los orígenes rudimentarios de las encuestas de opinión a las que nos hemos referido se encuentran ya en las antiguas Grecia y Roma, donde las esferas de poder contrataban a personas que anónimamente se mezclaban entre la población y pudiesen captar sus opiniones (Monzón Arribas, 1987:164)^[101]. Desde una perspectiva moderna, las encuestas se empezaron a realizar en varias revistas estadounidenses del s. XIX con motivo de las elecciones de 1825: las publicaciones aprovechaban sus páginas para realizar encuestas acerca de las intenciones de voto de sus lectores. (Soetzel, 1973)^[102]

Respecto a nuestro campo de estudio, situamos el año 1936 como el que alberga una novedad fundamental. Se considera un nuevo sistema de predicción y su nombre es “estudio de mercado”. Los primeros en utilizarlo serán Archibald Crossley, Elmo Roper y George Gallup. Su finalidad, en efecto, es la de preguntar abiertamente al consumidor acerca de sus necesidades, opiniones y deseos para poder así ofrecerle soluciones en un futuro. (Soetzel, 1973)^[102]

2.2.1.1.2. El cuestionario

El objetivo principal del cuestionario es conseguir que todos los entrevistados se encuentren en una misma situación que les permita entender de igual forma el tema a tratar cuando se les formulan las cuestiones. (Hague, 1994)^[103]

En función del tipo de respuesta que queramos obtener y del tipo de información que queremos alcanzar encontramos dos tipos de respuestas que se le pueden ofrecer a los entrevistados: las respuestas abiertas y las respuestas cerradas. (Berganza Conde, 2005)^[98]

Dentro de la tipología de preguntas, podríamos separarlas según su función. Podemos encontrar las “preguntas filtro”, usadas para eliminar a los entrevistados que no estén afectados por su temática. Existen las preguntas de control cuya finalidad es asegurar aspectos relacionados con la veracidad o la fiabilidad de determinados conocimientos o aptitudes. Las preguntas de introducción o “rompehielos”, también importantes, son las que se utilizan al inicio del cuestionario y que quieren ganarse la confianza de los entrevistados. (Sierra Bravo, 1982)^[104]

2.2.1.1.3. Análisis de contenido

Partiremos de las siguientes definiciones académicas para presentar el análisis de contenido. A dicha técnica de investigación se le atribuye objetividad y permite describir de forma sistemática y cuantitativa el contenido manifiesto de las comunicaciones (Berelson, 1952)^[105]. Entrando en detalle, el análisis de contenido será cualquier técnica de

investigación que sirva para hacer inferencias mediante la identificación sistemática y objetiva de características específicas dentro de un texto o en su contexto. (Holsti, 1969)^[106] (Krippendorff, 1990)^[107]

Los análisis de contenido sirven, como se ha apuntado, para obtener información del texto propiamente dicho, poniendo en relación algunas de sus partes o los conceptos elegidos y estudiarlos en función de su presencia e importancia respecto a la materia de nuestra investigación. Dentro del análisis de contenido deberemos matizar los diversos enfoques que han ido apareciendo con el desarrollo del método: cualitativo/cuantitativo, exploración/verificación/ manifiesto/latente. Dentro de las tipologías mencionadas nos interesa señalar la diferencia entre el enfoque cuantitativo y el cualitativo. Cada uno de los métodos presenta ventajas y desventajas y es preferido para un tipo de estudios u otros. El análisis cualitativo se centra en el hallazgo de temas y la descripción de particularidades y, con ello, establece categorías de análisis y ofrece diversas interpretaciones. En cuanto al análisis de contenido cuantitativo, destacaremos que los resultados obtenidos se basan en el recuento de las palabras, fragmentos de frases o sentencias elegidos. Los resultados obtenidos variarán en función de la frecuencia con la que aparezcan las unidades informativas que hayamos señalado. (Báez y Pérez de Tudela, 2009:290)^[108]

Como constata la siguiente afirmación: “En el ámbito de las ciencias sociales, entre las que se incluye la Comunicación, el “análisis de contenido” tiene un significado preciso y se enmarca habitualmente entre los métodos cuantitativos que se pueden emplear en el desarrollo de una investigación”. (Berganza Conde, 2005:207)^[98]

En nuestra tesis, como hemos ido desarrollando, partimos de la asunción de la gran utilidad en el campo de la comunicación de los análisis de contenido, sabiendo que son los que extraen información a partir de la medición de la frecuencia con la que aparecen ciertos términos en textos escritos. Vinculando directamente el análisis de contenidos a nuestro objeto de estudio, pensamos que las mencionadas conversaciones de los usuarios de Twitter que fluyen libremente acerca de un tema en concreto o un suceso en particular son susceptibles de ser analizadas textualmente. Su contenido, si logramos una selección de

palabras o fragmentos pertinentes y tenemos un tema de estudio acotado y claro puede ser de un valor incalculable para la obtención de información relevante acerca del consumidor y aquello que le rodea. Dentro de ello, de forma paralela, pensamos directamente en el valor que pueden tener ciertas informaciones bien organizadas y su potencial para permitirnos hallar nuevos *insights*.

2.2.1.1.4. Pre-test, post-test y estudios de seguimiento continuo

Dentro del ámbito del método cuantitativo de investigación, definiremos los estudios según pre-test, post-test y los estudios de continuidad. Como resulta evidente, esta tipología se debe al momento temporal concreto del proceso en el que se realiza el estudio.

Si bien es cierto que los estudios pre-test pueden ser tanto cuantitativos como cualitativos, los post-test son mayoritariamente cuantitativos. Según la investigadora Beriain, los siguientes datos ofrecen una panorámica acertada de cuáles de ellos son más utilizados. Del 100% del total de las herramientas orientadas a la medición cuantitativa de la eficacia de la comunicación publicitaria, su uso según el momento del proceso en el que se aplican responde a los siguientes porcentajes: (Beriain, 2011:398)^[109]

- 52,6% Pre-test
- 31,6% Estudio de seguimiento continuo
- 15,8% Post-test

Los estudios pre-test:

Los estudios pre-test que servirán al método cuantitativo pueden aplicarse dentro del marco de lo que hemos denominado métodos de investigación cualitativos, como las sesiones de *Focus Group* o las entrevistas individuales realizadas en profundidad.

En términos generales, el pre-test es la técnica de investigación publicitaria que consiste en escuchar y estudiar la opinión de una o varias personas que reflexionan y debaten entre sí la

campaña publicitaria que se les muestra. Esta acción se realiza de forma previa a la emisión o al inicio de la campaña o, como muy tarde, antes de su finalización¹³. El objetivo es verificar que alcanza las metas comunicativas que los publicistas tenían y que desde la óptica del marketing se cumplen también las expectativas. Además, permite detectar que no se produzcan reacciones adversas o imprevistas. (Alloza, 2000:856)^[111]

Hay varias formas de presentar un pre-test, de las que enumeramos las siguientes: *storyboard*, *animatic*, *replitomatic*, *low cost o spots*. Será gracias a los estudios pre-test que lograremos encauzar creativamente —si fuese necesario— una idea publicitaria y en especial podremos evitar posibles errores antes de la emisión de la campaña. (González, 1998:139)^[112]

Estudios de seguimiento continuo:

Los estudios de seguimiento continuo de la eficacia de la comunicación publicitaria o del estudio de la salud de una marca se utilizan principalmente en las campañas de larga duración.

En estos estudios pueden realizarse entrevistas diariamente y de forma continua a diversos segmentos del público objetivo, ya sea para recolectar datos sin interrupciones, analizar bloques de información o bien para elaborar medias móviles. (Alloza, 2000:847)^[111]

Como herramienta propia de los estudios de seguimiento continuo podemos observar, por ejemplo, la realización de un número idéntico de entrevistas a lo largo de los días laborables en periodos de cuatro semanas. En un ejemplo típico, con un tamaño muestral de trescientas entrevistas realizadas en total y en períodos de cuatro semanas, lo que suele hacerse es distribuir esas entrevistas entre los puntos de muestreo prefijados para el estudio y, de este modo, se volverá de manera rotativa a los mismos puntos en ese espacio global de tiempo”. (Alloza, 2000:847)^[111]

¹³ “Cualquier método científicamente fundamentado que permita determinar el grado en que un anuncio (sea cual sea su grado de finalización) logra los objetivos para los que fue creado, ofreciendo paralelamente posibilidades de optimización. Siempre se lleva a cabo antes del lanzamiento de la campaña publicitaria”. (Martínez, 2000:949)^[110]

Es interesante considerar que los estudios de seguimiento continuo son compatibles y pueden convivir con los estudios post-test puesto que ambos trabajan con el mismo objetivo: analizar la eficacia de las campañas una vez estas ya han concurrido. (Berriain, 2012)^[113]

Los estudios post-test:

Los estudios post-test suelen basarse en un cuestionario cerrado que dará la oportunidad a los encuestados para responder si la campaña ha cumplido o ha generado los efectos que los publicistas tenían previsto. En este aspecto se diferencia de los estudios pre-test ya que, al no saber a priori cuál será la reacción de los entrevistados, las preguntas deben ser abiertas y lo más libres posible sin escapar a los intereses del estudio. La siguiente definición sintetiza su valiosa utilidad: “los estudios post-test consisten en comprobar, una vez han tenido lugar todas las inserciones de que consta la campaña, que las personas del grupo objetivo han sido impactadas por los anuncios en la forma prevista y que el mensaje publicitario ha sido correctamente percibido”. (González, 1998:148)^[112]

2.2.1.2. Métodos cualitativos

Las técnicas de recogida de información que llamamos cualitativas engloban todas aquellas técnicas que pretenden encontrar el significado de la acción de los sujetos mediante la interpretación de la información que ellos mismos han ofrecido. En un método cualitativo pueden tomarse como valores tanto los sentimientos como las motivaciones de los individuos. (Berganza Conde, 2005:32)^[98]

Dentro de los métodos cualitativos de investigación se engloban técnicas tales como las entrevistas abiertas y en profundidad, la observación participante, los grupos de discusión y las historias de vida (García Ferrando, 1989)^[114]. Por el modo en que se aplican los métodos cualitativos, los insertamos dentro de la llamada lógica de la comprensión y entendemos que la interpretación subjetiva es un componente esencial para su desarrollo. (Schwartz, 1984:21)^[115]

2.2.1.2.1. La entrevista

La entrevista es la herramienta metodológica preferida para realizar investigaciones cualitativas (Denzin, 2011)^[116]. La llevan a cabo los investigadores de la escuela “abierta” y no la suelen utilizar los pertenecientes a la escuela “estructuralista”. En este sentido, desde la perspectiva constructivista, la entrevista no se considera un método neutral ya que intervienen demasiados factores en el proceso: el entrevistador, el entrevistado, la situación social en la que tiene lugar y finalmente la interpretación de la información obtenida. Sin embargo, desde una perspectiva menos estricta, se considera que, a pesar de estas limitaciones acerca de su neutralidad, los resultados obtenidos sí pueden ser relevantes. (Ruiz Olabuénaga, 1999:122)^[117]

Tipología de entrevistas:

- En profundidad, personales o cara a cara
- Telefónicas
- Por correo
- Por Internet

2.2.1.2.2. La entrevista en profundidad

Es una técnica de investigación para obtener información a través de una conversación profesional. Puede entrevistarse a una o a varias personas y gracias a la interpretación posterior del material obtenido se pueden realizar estudios analíticos de investigación, señalar diagnósticos o mostrar determinadas circunstancias sociales. En su vertiente más elaborada puede considerarse un arte: “la entrevista es fundamentalmente una conversación en la que y durante la que se ejercita el arte de formular preguntas y escuchar respuestas”. (Ruiz Olabuénaga, 1999:165)^[117]

La entrevista puede desarrollarse de formas diferentes: las más comunes serían las entrevistas habladas individuales o en grupo. A través de otros medios, como el correo o el teléfono,

encontraremos diversas formulaciones: estructuradas, controladas o libres. Obviamente, las entrevistas pueden tratar de un espectro muy amplio de temas, aunque pueden ser monotemáticas. Las entrevistas “estructuradas” hacen referencia a aquellas en las que el entrevistador lleva la iniciativa y marca las cuestiones y auspicia su desarrollo. En el caso de las “no estructuradas”, el entrevistador tiene un esquema general prediseñado pero lo desenvuelve de forma flexible a través de las preguntas y cierta improvisación. (Ruiz Olabuénaga, 1999)^[117]

2.2.1.2.3. La entrevista telefónica

Por su facilidad y rapidez a la hora de ponerlas en marcha, así como por su menor coste en comparación con las entrevistas personales, es una de las tipologías de entrevista más utilizadas. (AEDEMO, 2000:44-45)^[118]

2.2.1.2.4. La entrevista por correo

Representan un método sencillo y económico y es especialmente útil para investigar muestras muy dispersas. Permiten una mayor extensión en las preguntas y también una mayor obtención de respuestas. Los principales problemas que presentan son las dudas acerca de si será posible alcanzar el público realmente deseado y si la tasa de respuesta será suficiente (suele situarse entre el 60% y el 80%). (Fox, 1988:468)^[119]

2.2.1.2.5. La entrevista por Internet

Los primeros programas por ordenador utilizados para emprender análisis cualitativo de información se sitúan a principios de la década de 1980 (CAQDAS: *Computer-Assisted Qualitative Data Analysis Software*, “*The Ethnograph*”, QUALPRO, QUALOG y Notebook). Con el paso de los años se han introducido numerosas mejoras y cambios en los programas que todavía siguen utilizándose. (Weitzman, 2000:804)^[120]

Como es evidente, Internet ofrece la posibilidad de realizar todos los tipos de entrevistas mencionados anteriormente pero a través de programas desarrollados en un sitio web y sin la necesidad de un entrevistador, un teleoperador/a o el tener que recurrir a los servicios de Correos.

Las denominaciones más comunes de los programas son: (Johnson, 2008)^[121]

- CAPI (*Computer-assisted personal interviewing*)
 - Entrevista personal asistida por ordenador
- CATI (*Computer-assisted telephones interviewing*)
 - Entrevista telefónica asistida por ordenador
- CAWI (*Computer Aided Web Interviewing*)
 - Entrevista asistida por ordenador

Enumeramos brevemente las ventajas e inconvenientes de las entrevistas realizadas por Internet:

Ventajas:

- Menor influencia tanto del investigador como del entorno. Permiten mayor comodidad y que el entrevistado se exprese de forma más sincera.
- Mayor flexibilidad del administrador a la hora de presentar las preguntas y la forma de responderlas: botones, cajas de marcación, menús desplegables. (Preece, 2015)^[122]
- Mayor facilidad, rapidez y solución de errores en la formulación de la entrevista, dado que no es necesaria su reimpresión.
- Mayor rapidez en la recepción de respuestas. (Bradburn, 2004)^[123]
- Reducción del tiempo de análisis gracias a la accesibilidad a las bases de datos.

Inconvenientes: (Vilches, 2011:215)^[124]

- Reducción de la espontaneidad
- Menor interacción y menor presencia del aspecto emocional
- Dinámica e interacción más pobre que con otros métodos
- Riesgo de posibilidad de suplantación

2.2.1.2.6. Focus Group o “Grupo Focal”

El término *Focus Group* se atribuye al experto en marketing Ernest Dichter¹⁴ (Goebert, 2002:214)^[125]. Se conoce también como Grupo Focal o Grupo de Discusión (Gutiérrez Brito, 2011)^[126]. Históricamente, se sitúa el primer *Focus Group* en el “*Bureau of Applied Social Research*” en la Universidad de Columbia en Estados Unidos y lo llevó a cabo el sociólogo Robert K. Merton en el año 1946. (Kamberelis, 2013:3)^[127]

El número de participantes más habitual en los *Focus Group* puede oscilar entre los seis y los doce participantes, siendo su número ideal inferior a diez (Liamputtong, 2011:42)^[128]. La selección de los participantes depende del tema a tratar y su desarrollo será dirigido por el entrevistador que coordinará al grupo para asegurar que se alcanzan los objetivos del estudio.

Desde la óptica del marketing, Mucchiell aporta una definición centrada en su utilidad en ese campo: “Por grupo de discusión, en sentido estricto, se entiende un grupo cuyo objetivo es llevar a cabo una confrontación de opiniones, de ideas o de sentimientos de los participantes, con vistas a llegar a unas conclusiones, a un acuerdo o a unas decisiones”. (Mucchielli, 1974:107)^[129]

El investigador Ibañez, ofrece una amplia y detallada definición sobre Grupo de discusión que transcribimos a continuación: “Un grupo de discusión es un dispositivo analizador cuyo proceso de producción es la puesta de colisión de los diferentes discursos y cuyo producto es

¹⁴ Ernest Richter (Viena 1907 - Nueva York 1991). Psicólogo y experto en Marketing conocido como el “padre de la investigación motivacional”.

la puesta de manifiesto de los efectos de la colisión (discusión) en los discursos personales (convencimiento: convencido es el que ha sido vencido por el grupo) y en los discursos grupales (consenso)”. (Ibáñez, 2000:58)^[130]

Recogemos aquí otra definición de *Focus Group* basada en los textos que ofrece Callejo: técnica cualitativa de estudio de opiniones o actitudes de los consumidores y que consiste en la reunión de un grupo de entre seis y diez personas, desconocidas entre sí, dirigidas por un moderador o investigador, encargado de realizar las preguntas y controlar la discusión del grupo para que alcance los objetivos deseados. (Ortí, 1989:198)^[131] (Morgan, 1991:13)^[132] (Callejo, 2001:25)^[133]

Cerraremos el apartado de los *Focus Group* con otro apunte realizado por Percy y Moriarty que amplía y matiza lo que ya hemos apuntado anteriormente:

Focus groups are used with surprising frequency for making final go/no-go decisions. Reasons usually given are: speed of feedback and reduced cost, despite evidence (and exhortations within the research industry) against the practice. (Percy, 1997)^[134] (Moriarty, 2014)^[135]

2.2.1.2.7. El método Delphi

Como método: “método de estructuración de un proceso de comunicación grupal que es efectivo a la hora de permitir a un grupo de individuos tratar un problema complejo funcionando como un todo”. (Linstone, 1975:3)^[136]

Se atribuye el nombre del método al filósofo Abraham Kaplan; en la década de 1940 formó parte de un grupo de investigadores en búsqueda de una alternativa a la metodología positivista y que pudiese satisfacer las demandas procedentes del campo de las disciplinas sociales (Landeta Rodríguez, 2002:31-35, 93-94)^[137]. En la década siguiente, el método Delphi sería desarrollado por Olaf Helmer, Nicholas Rescher, Norman Dalkey y otros miembros de la *Rand Corporation*. (Gordon, 1994)^[138]

Características y requisitos básicos del método Delphi:

1. Anonimato: Ningún miembro del grupo conoce a quién corresponde una respuesta en particular. Entre las ventajas del anonimato se encuentra el poder evitar las influencias negativas de los miembros dominantes del grupo o la inhibición de algún participante. El control de la comunicación sigue en manos del grupo coordinador y nunca se establece una participación directa entre los expertos involucrados.

2. Realimentación o *feedback* controlado: al inicio de cada ronda, el grupo coordinador pone en común las posiciones que los participantes han mostrado anteriormente y analiza la situación destacando las aportaciones más significativas.

3. Respuesta estadística del grupo: en el caso de que al grupo se le haya solicitado una estimación numérica, se manejará la mediana de las respuestas individuales. De este modo, se consigue la inclusión de las respuestas individuales en el resultado final del grupo. (Dalkey, 1969:16)^[139]

2.2.1.3. Cuadro de resumen de los métodos tradicionales de investigación

A continuación desarrollamos un cuadro resumen de los métodos tradicionales de investigación donde destacamos aquellos aspectos más relevantes para su mejor comprensión y comparación con los métodos de investigación que desarrollaremos a continuación.

Método	Proceso	Tipología	Objetivos
Entrevista (individual ó telefónica)	Preguntas a los entrevistados sobre hechos, opiniones y actitudes	Cualitativa	Identificar factores que influyen en la compra de un producto o servicio
Encuestas (online, correo, <i>in-store</i>)	Preguntas a los participantes o	Cuantitativa y/o Cualitativa	Ej.: ¿Cuál es el precio más

	selección de respuestas del cuestionario.		adecuado para vender un producto o servicio?
<i>Focus Groups</i> (selección de consumidores)	El entrevistador interactúa con el grupo para obtener opiniones y actitudes	Cualitativa	Ej.: ¿Qué áreas serían susceptibles de mejora?
Cuestionarios: (online, correo, in-store)	Los participantes seleccionan las respuestas predeterminadas que consideran más adecuadas	Cuantitativa	Ej.: Puntuación de productos

Tabla 2: Cuadro resumen de los métodos tradicionales de investigación. (Straker, 2013:5048)^[140]

2.2.1.4. Conclusión acerca de los métodos de investigación tradicionales

En los siguientes dos apartados señalamos tanto los aspectos positivos que tiene aplicar un determinado método de investigación tradicional como las limitaciones que presenta su desarrollo. Si mostramos las ventajas y los inconvenientes de los métodos tradicionales podremos apreciar qué nuevas vías se han abierto con la aparición de los medios sociales y detectar, de este modo, el añadido que pueden ofrecer en cuanto a la obtención de *insights* publicitarios.

Ventajas de los métodos tradicionales de investigación:

Según se ha demostrado, los métodos de investigación tradicionales funcionan apropiadamente cuando se dan ciertas circunstancias: pequeños cambios en el público objetivo, en el entorno competitivo o también cuando se puede acceder con facilidad al recuerdo que necesitamos de los clientes y a sus necesidades respecto a nuestros intereses. (Zaltman, 2003)^[141]

Consideramos otro aspecto positivo de los métodos tradicionales de investigación el hecho de que los *insights* que se pueden llegar a obtener permiten identificar claramente las jerarquías de productos o los mejores precios para un producto o servicio. (Sen, 2009)^[142]

Asimismo, como apunta Zaltman, las encuestas y cuestionarios facilitan la labor de análisis estadísticos y permiten gestionar grandes cantidades de datos de forma rápida. Las encuestas presentan el añadido de permitirnos identificar correlaciones entre respuestas que podrían no aparecer en otros métodos. (Zaltman, 2003)^[141]

En el aspecto económico, las encuestas desarrolladas a través de Internet son capaces de alcanzar públicos objetivos más amplios con presupuestos más económicos. (Sen, 2009)^[142]

Se reconoce la fiabilidad y utilidad de las encuestas para la obtención de informaciones relativas a cuestiones familiares, identificación de tomas de decisión inminentes y detección de necesidades. (Griffin, 1993)^[143]

Como ventaja añadida, es preciso mencionar que los *Focus Group* son actualmente fáciles de implantar y se pueden llevar a cabo con un coste reducido.

Inconvenientes de los métodos tradicionales de investigación:

- La mayoría de métodos de investigación son poco accesibles a la pequeña y mediana empresa, tanto por la experiencia que es necesario poner en práctica como por los presupuestos limitados que pueden dedicar a llevarlos a cabo. (Sen, 2009)^[142]
- Los sujetos que participan en la investigación llevan a cabo un ejercicio de auto-reflexión y de recuerdo de las marcas que han consumido previamente. Al ser preguntado explícitamente por ello, al consumidor le resulta casi imposible evitar el darle mayor importancia a los recuerdos de la que les ha dado anteriormente (Zaltman, 2003)^[141] (Graves, 2010:4)^[144]. Con ello queremos decir que dichos métodos generan principalmente una actitud reactiva y retrospectiva ya que las propias preguntas y su formulación limitan las mismas respuestas. (Witell, 2011)^[145]

- Los *Focus Group* son incapaces de proporcionar las condiciones totalmente óptimas para que los participantes puedan profundizar en las discusiones. Si bien los entrevistadores pueden intentar generar el mejor escenario posible, este nunca será el más espontáneo ni el que permita a los entrevistados mostrar con plena naturalidad sus pensamientos y sentimientos. (Zaltman, 2003)^[141]

Existe otra dificultad muy comentada y que hemos sugerido anteriormente: inevitablemente, la presencia del entrevistado o de la existencia de ciertas preguntas elegidas condicionan o influyen, en cierto modo, las mismas respuestas de los entrevistados (Zaltman, 2003)^[141]. Y, aunque la pregunta esté perfectamente formulada, en muchas ocasiones los entrevistados no saben expresarse del modo más adecuado en función del tipo de entrevistados que tienen, y puede que permanezca ajeno a las necesidades reales de los presentes (Goffin, 2012)^[146]. En este sentido, puede que los propios entrevistados respondan a ciertos temas o a productos nuevos saliendo del tema adecuado para el análisis, especialmente si carecen de experiencia previa sobre lo que se les pregunta o sugiere. (Goffin, 2012)^[146] (Trott, 2001)^[147] (Witell, 2011)^[145]

2.2.2. Técnicas proyectivas de investigación

Consideramos necesario incluir las técnicas proyectivas de investigación por el creciente interés que han tenido dentro de los métodos de investigación cualitativa en los últimos años (Fram, 1991)^[148] (Miller, 1991)^[149] (Catterall, 1998)^[150]. Dentro de esta categoría, destacamos el especial dinamismo de la técnicas productivas (Miller, 1991)^[149]. Las técnicas productivas buscan nuevas vías para alcanzar un nivel más profundo de conocimiento de los factores que motivan a los consumidores.

Remontándonos a sus orígenes, dichas técnicas se empezaron a utilizar a partir de la década de 1940 (Catterall, 1998)^[150] para conseguir que los entrevistados expresaran sus sentimientos y actitudes que con técnicas de investigación más directas solían mantener ocultas por vergüenza o miedo. Las disciplinas que han dado lugar a su nacimiento son el Psicoanálisis y la Psicología Clínica (Robson, 2000)^[151] (Boddy, 2004^a)^[152] donde se siguen utilizando en la actualidad. (Richman, 1996)^[153]

La denominación “Proyección” se atribuye a Sigmund Freud y a sus estudios sobre la paranoia. Él conceptualizó la proyección como un mecanismo de defensa según el que la gente, de forma inconsciente, proyecta sus rasgos de personalidad negativos en las otras personas. (Lilienfeld, 2000)^[154]

El Diccionario de Psicología de la *Oxford University Press* define las técnicas proyectivas como:

Any of a variety of personality tests in which the respondent gives free responses to a series of stimuli such as inkblots, pictures, or incomplete sentences. Such tests are based loosely on the psychoanalytic concept of projection, the assumption being that respondents project unconscious aspects of their personalities on to the test items and reveal them in their responses. (Colman, 2015:606)^[155]

Las técnicas proyectivas se utilizan principalmente en estudios de mercado cualitativos (Gordon, 1988)^[138] aunque también se utilizan para estudios cuantitativos y pueden ser útiles para ambos estudios. (Boddy, 2004^b)^[152]

El objetivo de las técnicas proyectivas es conseguir que el entrevistado se proyecte a sí mismo dentro de la situación creada y verbalice sus pensamientos sugeridos a raíz de dicha situación. Existen una gran variedad de técnicas proyectivas, tales como la asociación de palabras, completar frases o completar historias. Son técnicas eficaces cuando hay aspectos de la investigación que no pueden ser alcanzados mediante las preguntas directas. (McDaniel, 1995)^[156]

Como todas las técnicas de investigación cualitativa, las críticas que reciben cuestionan su fiabilidad: al existir una interpretación que depende del entrevistador existe una controversia alrededor del valor científico de los resultados obtenidos. (Boddy, 2004^a)^[152]

Presentamos a continuación una tabla que expone algunas de las técnicas proyectivas más recientes junto a una breve descripción:

Método	Proceso	Tipología	Objetivos
<i>Persona Design</i> ¹⁵ (individual)	Se crea un perfil de cliente con máximo detalle: demográfico, comportamiento, objetivos,...	Cualitativa	El perfil sirve para acercarse al cliente en busca de respuestas sinceras
<i>Story-telling/Story Board</i> (individual)	Se pide al usuario/cliente que cuente una historia a partir de su experiencia	Cualitativa	Permite identificar contradicciones o errores con el uso del producto
<i>Customer Narratives</i> (individual)	Se analizan en profundidad todos los puntos de	Cualitativa	Un discurso del cliente que explore su experiencia

¹⁵ Definición de “Persona”: en el diseño centrado en el usuario, “persona” es un personaje ficticio creado para representar los diferentes tipos de usuarios que pueden utilizar un sitio web, un producto o una marca, de forma similar (Lidwell, 2003:182)^[157].

	contacto entre el consumidor y las empresas, productos o servicios		como comprador
<i>Scenarios (individual)</i>	Se crea un guión a partir de imágenes y breves descripciones que contextualicen la historia	Cualitativa	Ilustra diferentes escenarios de uso actual para diseñar escenarios futuros.

Tabla 3: Tabla resumen de las técnicas proyectivas más recientes. (Straker, 2013:5050)^[140]

2.2.3. Técnicas específicas de generación de *insights*

Las necesidades específicas de la búsqueda de *insights* han llevado a los anunciantes y las agencias a desarrollar sus propias técnicas, más allá de quedarse en una metodología de uso general.

Como veremos a continuación, todas coinciden en algunos puntos necesarios dentro de un método de investigación: la búsqueda de información, profundizar en el contenido de la información con el fin de identificar conexiones, agrupar y ordenar la información, el análisis en profundidad para detectar oportunidades y nuevas formas de solucionar un problema o satisfacer una necesidad.

La etapa final del proceso es la que puede resultar más confusa respecto al hallazgo o no de un *insight* tal y como lo hemos descrito previamente. Para evitarlo, Hartman desarrolla una lista de cinco puntos clarificadores: (Hartmann, 2007:112)^[84]

1. Conduce a la acción. El *insight* tiene el poder de activar el pensamiento creativo y debe ser capaz de estimular la mente de los miembros del equipo de trabajo.
2. Conduce a conectar con otras personas; es capaz de establecer cierta interacción con los demás. En cierto modo, podríamos decir que es algo que forma parte de la misma condición humana.
3. Conduce a mejorar en algún aspecto la vida de las personas.
4. Tiene una razón de ser fundamental, básica y de primer orden.
5. Posee un carácter universal y no tiene un dominio geográfico limitado.

Si continuamos trabajando con estudios que nos permitan definir mejor el punto en el que emerge el *insight*, encontramos la afirmación de la investigadora Williams: los *insights* son hallazgos, pero no todos los hallazgos son *insights*. La investigadora propone una metodología que se compone de tres aspectos útiles y que debemos tener en cuenta para entender con mayor profundidad y precisión si hablamos o no de un *insight*:

- a. El contexto que rodea los hallazgos
- b. Las implicaciones en el comportamiento de los consumidores respecto al hallazgo
- c. Cómo se aplica dicho descubrimiento a los objetivos del marketing

Las investigadoras de TNS¹⁶ India, Shashikala Raj y Anjali Puri, expusieron su visión del proceso de generación de *insights* del modo siguiente: “The structure of a detective model can be broken into five different stages:” (Srivatsa, 2007:2)^[158]

1. *The problem*
2. *The data gathering or getting down and dirty*
3. *Observation and the assimilation of atmosphere*
4. *The analysis or the piecing together of a jigsaw*
5. *Imagination and the balance of probabilities.*

Los investigadores Greg Rowland, director de la consultoría *Greg Rowland Semiotics*, y Jaroslav Cír, representante de Unilever, expusieron un ejemplo de generación de *insights* llevado a cabo en Rusia basándose en el caso de la marca de desodorantes Rexona.

Rowland y Cír quisieron explicar a través de la semiótica y el uso de estudios etnográficos cómo descubrir los *insights* en una cultura tan hermética como la rusa. A partir del insight detectado, mostraron cómo pudieron construir una plataforma estratégica desde la que la marca pudiese operar y anclar su territorio de comunicación.

La semiótica les permitió entender las complejidades y paradojas de la cultura rusa, así como establecer las relaciones simbólicas entre la marca y el consumidor. (Rowland, 2007:37)^[159]

La relación de técnicas que se exponen a continuación se han obtenido a partir de la tesis doctoral de Cristina de Balanzó sobre los *Consumer insights* y las neurociencias fruto de las

¹⁶ Taylor Nelson Sofres TNS. Ahora pertenece al grupo Kantar Media, líder mundial en conocimiento y comprensión del consumidor a través de paneles de consumo. (Kantar, 2015)^[160]

entrevistas realizadas a los responsables de las empresas y agencias que se mencionan a continuación para conocer sus técnicas de elaboración de *insights*.

Nestlé, como gran fabricante y anunciante que es de productos de gran consumo, ha tenido y sigue teniendo muy en cuenta las informaciones procedentes de sus consumidores como fuente relevante de información para la mejora de sus estrategias de marketing y de comunicación. En este sentido Nestlé ha elaborado su propia metodología para la obtención de *insights* que a continuación presentamos y que se basa en los siguientes momentos clave:

1. “El qué”: obtención y análisis de datos e información.
2. “El cómo”: buscar, a partir de la información, los lazos, palabras y ramificaciones comunes. Con ello se hace una lista de entendimientos principales.
3. “¡Ajá!": el *insight*. El momento del descubrimiento.

Para dar por válido un *insight*, éste debe estar formado por las ideas más fuertes y relevantes y que en cuyo proceso han necesitado de “ayudas” para convertir este conocimiento en *insights* significativos y que describe como los “puntos calientes” o “palabras mágicas” : sorpresas, contradicciones, emociones, palabras utilizadas con mayor frecuencia, dicho con emoción o convicción, entre otras. Otra ayuda la considera la comparativa entre el: “Ahora frente a lo Nuevo” y que refiere a la confirmación de la existencia de conocimiento, la obtención de nuevos aprendizajes y temas a cambiar y/o construir y finalmente, Nestlé se apoya sobre una tercera “ayuda” que estaría formada por las “Barreras” al consumo que el nuevo *insight* podría solucionar.

Por su lado Unilever desarrolla su metodología en cuatro fases:

1. Definición de la problemática sobre la que debemos actuar
2. Búsqueda y recolección individual de observaciones a partir de la revisión de estudios realizados, de reuniones con consumidores, de artículos, experiencias o pensamientos.
3. En esta fase se agrupan las observaciones en plataformas: formas de consumo, *target*, motivos de compra o tendencias, por ejemplo.
4. La fase final es la que constituye propiamente la construcción de *insights*.

Por su lado la agencia Salvetti-Llompart desarrolla una metodología denominada “Las 3 ‘i’ del *consumer insight*”

1. Información: análisis de toda la información disponible
2. Inmersión: observación del consumidor en su entorno natural y realización de una entrevista. Se trata de observar, descubrir y comprender las necesidades, motivaciones, deseos, instintos e identidades inconscientes que luego se ubicarán en un mercado, una situación y un producto o servicio.
3. Intuición: a partir de la información obtenida, se procede a la segmentación de los *insights* agrupándolos por mercados, productos y situación.

Finalmente, la metodología desarrollada por el Instituto Quid, se basa en la elaboración de un espacio que tiene en cuenta dos aspectos importantes, por un lado, la relación entre la marca y el consumidor y por otro que el *insight* se halla oculto dentro de su discurso racional.

Con estas premisas, Quid desarrolla las siguientes fases para la elaboración de *insights* y que empiezan por:

1. La preparación: puesta en común de los objetivos que se desean alcanzar. En esta fase participan el anunciante, la agencia de publicidad y el instituto de investigación.
2. Extracción de *insights*: se trata de romper con el bloqueo del consumidor para conseguir establecer la conexión con él.
3. “Operativización” de los *insights*: adaptación del *consumer insight* a la marca de modo que pueda tener para su desarrollo una nueva utilidad.

Sin lugar a dudas, todos los métodos que acabamos de presentar son legítimos y ha quedado demostrado que a las tres empresas les ha surtido un efecto positivo. Podríamos afirmar que lo más importante son la intuición y la perspicacia para estar atentos a todas las partes del proceso y a los resultados de las herramientas que hemos utilizado. Como hemos ido desarrollando, el paso necesario para que los métodos de investigación sean útiles para

nuestro cometido es conseguir dar el paso de los datos a los *insights* y que alcancen la verdadera forma de ser, sentir y pensar del consumidor. (Balanzó, 2011:104)^[86]

A modo de resumen, en este punto del capítulo hemos procurado trazar un recorrido a través de los medios de comunicación que nos permitiese llegar hasta la concreción de los medios sociales que son el centro de esta tesis. Era primordial mostrar la diferencia entre los medios de comunicación masivos “tradicionales” y los medios de comunicación sociales. Son los Medios Sociales los que han permitido desarrollar nuevos métodos de investigación que las empresas han aplicado, además de en otros campos, en la publicidad. Para ello, las tablas comparativas nos han permitido recalcar las diferencias y similitudes entre los tipos de métodos y hemos trabajado con la explicación de los métodos cuantitativos y los cualitativos. Llegados a ese punto, nuestro objetivo ha sido señalar aquellos métodos que permiten la obtención de *insights*. A modo ilustrativo y con el fin de comenzar a estudiar sobre casos reales, hemos hablado de los métodos de obtención de *insights* que han desarrollado ex profeso distintas empresas en varios sectores. Con este marco que ha pasado de lo general a lo concreto, podemos ahora adentrarnos en la explicación y el desarrollo de la Teoría de Grafos, herramienta esencial para la obtención de *insights* que aspiramos a proponer en los capítulos siguientes y que tiene las conversaciones de Twitter como fuente principal de estudio.

3. TEORÍA DE GRAFOS

3. Teoría de Grafos

La Teoría de Grafos, en especial los avances y evoluciones que ha ido recibiendo a lo largo de las últimas décadas de la mano de diversos investigadores y académicos, se ha postulado como una herramienta indispensable para el análisis de la actividad en las relaciones sociales que tienen lugar entre sus usuarios. Para nuestra investigación es la herramienta que más información relevante y de calidad podrá reportarnos.

En este apartado hemos querido desplegar de la forma más completa posible sus usos, tanto actuales como iniciáticos, con el fin de obtener una comprensión global de sus planteamientos. Trazaremos un recorrido histórico desde sus inicios hasta los avances recientes y dedicaremos un apartado especial a su aplicación en cuanto al estudio y análisis de las redes sociales (ARS).

Dentro del apartado dedicado al ARS, desplegaremos las combinaciones que pueden darse con otras herramientas para perfilar nuevos hallazgos y alcanzar informaciones relevantes en cuanto a las acciones de sus usuarios: actitudes, vínculos o la capacidad de influencia (los usuarios llamados *influencers*) y el acceso a la información de los usuarios serán, entre otros datos, material que obtendremos para realizar nuestro estudio vinculado al descubrimiento de *insights*.

3.1. Antecedentes

Con el fin de comprender en profundidad la Teoría de Grafos, pilar fundamental para el desarrollo de nuestro estudio en los próximos capítulos, hemos trazado una retrospectiva de su evolución. Por su importancia en multitud de campos del saber y la investigación, son numerosos los teóricos y académicos que han realizado aportaciones de enorme valor y que queremos detallar a continuación. Para segmentar y organizar cronológicamente las aportaciones, hemos optado por separar el desarrollo histórico según los años en los que la Teoría de Grafos recibió anotaciones, cambios o mejoras.

1736:

Fue el planteamiento de un problema en concreto el que llevó a Leonhard Euler¹⁷ a presentar el concepto de grafos que, posteriormente, daría paso al desarrollo de la Teoría de Grafos. Es el llamado “Problema de los puentes de Königsberg” (hoy Kaliningrado). (Biggs, 1986:128)^[161]

El problema del puente de Königsberg es quizás el ejemplo más conocido de la Teoría de Grafos; un problema con un largo período sin solución, hasta que lo resolvió Leonhard Euler a partir de un grafo (Deo, 2004:3)^[162]. La ciudad de Kaliningrado es famosa por sus siete puentes que unen ambos márgenes del río Pregel con dos de sus islas, de modo que la ciudad quedaba dividida por el río en cuatro regiones: dos orillas, una sola isla entre ellas y un segmento de tierra. Las orillas B y C se encontraban unidas con la isla A y con el segmento de tierra D por medio de siete puentes (Biggs, 1986)^[161]. El problema que se planteaban era cómo regresar, partiendo de un punto cualquiera, al lugar de partida cruzando cada puente, una sola vez. (Foulds, 1992)^[163]

Euler fue el primero en plantear este problema enfocándolo desde un punto de vista matemático, y asignó a cada isla un nodo (vértice) y a cada puente un enlace (línea o *edge*).

¹⁷ EULER, Leonhard. “*Commentarii Academiae Scientiarum Imperialis Petropolitanae*” (1736), 128-140. Título del libro original.

Euler sustituyó cada una de las cuatro regiones (A, B, C y D) por un punto (nodo) y unió los puntos con líneas (aristas), produciendo así el primer grafo.

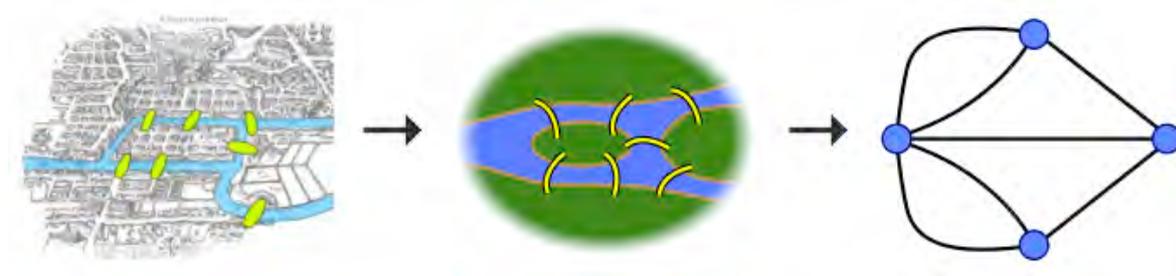


Figura 1: ilustración que representan la ciudad de Kaliningrado, el río Pregel y sus siete puentes que fueron el origen del primer grafo. (Van Dooren, 2009)^[164]

Como se puede observar en las imágenes, no es posible atravesar cada una de las líneas del grafo sin repetir por lo menos una de ellas. El hecho de que haya más de dos vértices de grado impar hace el resultado satisfactorio imposible. Por contra, si dos vértices tuviesen grado impar, sí sería posible hallar solución al problema planteado. Finalmente, si no hubiese vértices de grado impar, el problema se podría resolver empezando desde cualquiera de ellos. (Hopkins, 2005:206)^[165]

1802 - 1830:

En el año 1802, el etólogo Pierre Huber observó las relaciones de dominio entre las abejas en sus colmenas y realizó el que sería el primer estudio de una de las líneas más amplias de estudio de la Teoría de Grafos y de la Teoría de Redes Sociales: las jerarquías y las dominancias. (Scott, 2011:34)^[166]

Auguste Comte¹⁸ realizó, en el año 1830, una de las primeras y más destacadas afirmaciones explícitas desde un enfoque estructural de la vida social. Tan es así que a él se le atribuye el primer uso del término “sociología”. (Estrada, 2011)^[167]

¹⁸ Isidore Auguste Comte (Montpellier. Francia. 1.798-1.857) Creador del positivismo y de la sociología.

1847:

Gustav Kirchhoff¹⁹ aplicó la Teoría de Grafos al análisis de las redes eléctricas. En sus estudios, consideró que una red eléctrica es un grafo lineal orientado y que cada uno de sus arcos lleva asociado dos funciones del tiempo (t), la intensidad de corriente (i) y la tensión (v). Posteriormente, Kirchhoff publicó sus “Leyes de Kirchhoff” para el cálculo del voltaje y la corriente de los circuitos eléctricos y anunció dos reglas que se consideran la base del análisis del flujo de corriente eléctrica en una red de cables, todavía vigentes en la actualidad (Foulds, 1992)^[163]. La singular aplicación de la Teoría de Grafos de Kirchhoff a la ingeniería es vista como pionera en este campo (Balabanian, 1972:87)^[168]. Con las Leyes de Kirchhoff, el investigador reemplazó cada red por su respectivo grafo asociado y demostrando que no era necesario considerar cada ciclo de un grafo separadamente para resolver el sistema de ecuaciones. Por el contrario, demostró que los ciclos independientes de un grafo, determinados por cualquiera de sus árboles generadores, son suficientes para resolver un sistema de ecuaciones. (Balabanian, 1972:41)^[168]

1852:

Dentro de los primeros estudios que hemos apuntado acerca de la aplicación de la Teoría de Grafos en diversos campos, uno de los más célebres y productivos se presentó en 1852 y hace referencia al “Problema de los cuatro colores”.

Las primeras referencias sobre dicho estudio se sitúan en el año 1840 y las llevó a cabo Möbius²⁰ pero sería en 1852 que Francis Guthrie²¹ trabajaría con mayor exactitud el caso. Guthrie observó que era posible pintar un mapa con solamente cuatro colores, siempre manteniendo los países con fronteras colindantes pintados de diferente color. El problema lo difundiría Cayley en 1879 con la publicación del primer volumen de “*Proceedings of the Royal Geographic Society*”. (Deo, 2004:10)^[162]

¹⁹ Gustav Kirchhoff (Königsberg 1824, Berlín 1887) físico prusiano. Destacado científico en el campo de los circuitos eléctricos, óptica, espectroscopia y radiación.

²⁰ August F. Möbius (1.790-1.868) fue un matemático alemán. Descubridor de la banda de Möbius.

²¹ Francis Guthrie. (Londres 1831 – Claremont Cap 1899) Matemático y Botánico sudafricano. Creador del Teorema de los cuatro colores.

Esta fue la cuestión que a día de hoy sigue citándose en el campo académico:

Is it true that any map drawn in the plane may have its regions colored with four colors, in such a way that any two regions having a common border have different colors?. (Fritsch, 1998:2) ^[169]

En efecto, por complejo que sea un mapa, este se puede pintar usando un máximo de cuatro colores.

Desde la Teoría de Grafos “El problema de los cuatro colores” se plantea de la siguiente manera: cada país se convierte en un vértice del grafo y la relación R entre sus vértices viene dada por los países que tienen una frontera en común. Con este procedimiento se obtiene así un grafo que representa fielmente las relaciones de vecindad de los países del mapa. A dicho grafo se le llama grafo planar. Una vez reformulados los mapas como grafos planares, el problema de los cuatro colores se reduce a un problema entre equivalencias de grafos. La solución al problema de los cuatro colores responde a la pregunta de si es posible pintar los vértices de cualquier grafo planar con cuatro colores o incluso menos. La respuesta es afirmativa. (Gonthier, 2008)^[170]

Aunque en términos del avance en la aplicación de la Teoría de Grafos no sea relevante, sí resulta sorprendente que la demostración matemática de la conjetura de los cuatro colores se logró finalmente en 1976 y utilizando un ordenador para realizar los cálculos necesarios. (Appel, 1976)^[171]

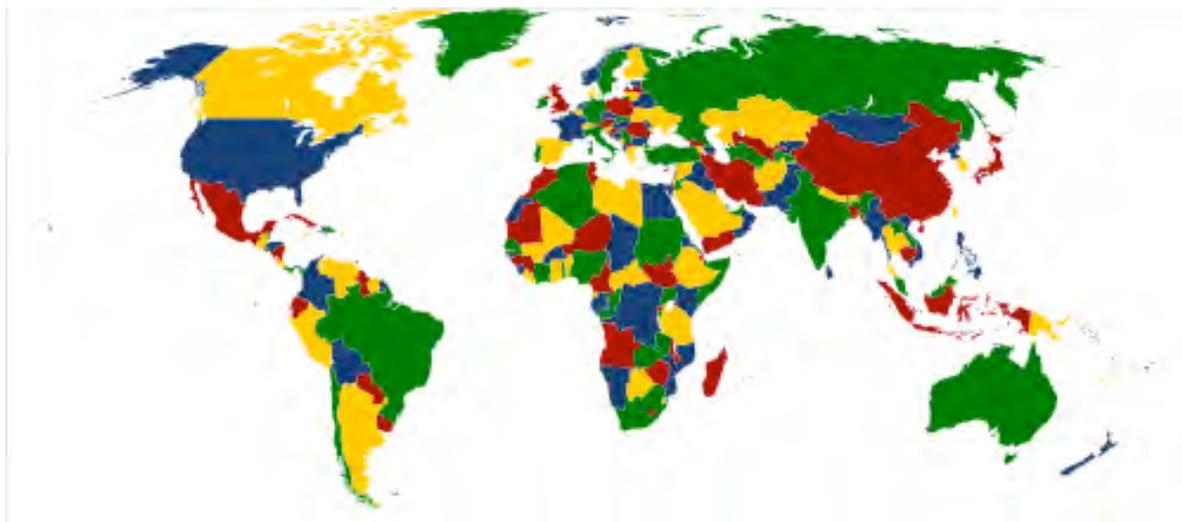


Figura 2: Ejemplo de uso y representación sobre el mapa del mundo del Teorema de los cuatro colores. (Wikipedia, 2014)^[172]

1856:

El planteamiento que realizó en 1856 Sir William Rowan Hamilton²² consistió en el uso del sólido regular de doce pentágonos (dodecaedro) y en ubicar etiquetas en cada uno de sus veinte nodos con nombres de ciudades famosas. Nació así el juego de “Alrededor del mundo”, consistente en encontrar un camino que pueda recorrer todos los vértices a través de sus líneas y de una sola vez (Chartrand, 1977)^[173]. El grafo conocido como “Grafo del dodecaedro” es uno de los cinco grafos platónicos existentes y surge al etiquetar los vértices del grafo con los números del 1 al 20, siendo cada uno de ellos una ciudad.

²² William Rowan Hamilton (1805 – 1865). Matemático, Físico y Astrónomo irlandés.

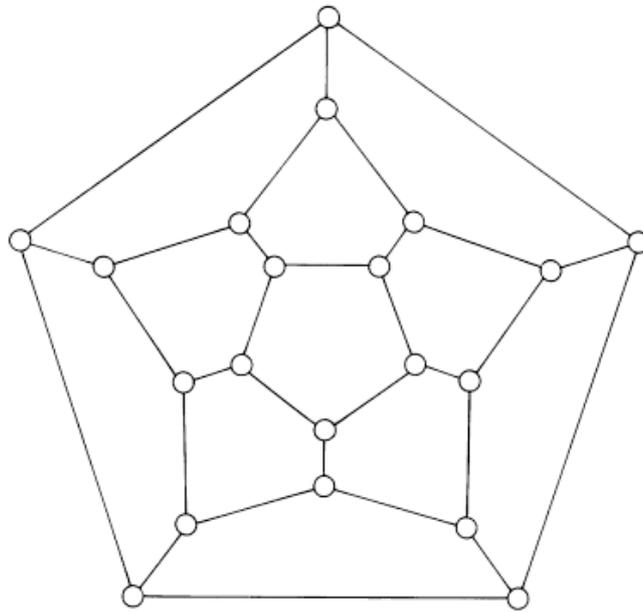


Figura 3: Representación gráfica del juego de Hamilton. (Foulds, 1992:XXII)^[163]

1857:

Se atribuye a Arthur Cayley, profesor de la *British Mathematician*, el uso por vez primera del término “árbol” (*tree*) para definir a un grafo sencillo no dirigido G . Sería en el año 1857 cuando Cayley denominaría árbol al par de vértices que quedan conectados por exactamente un único camino. En otras palabras, cualquier grafo conectado sin círculos constituye un árbol. (Crilly, 2006:44)^[174]

1878:

En cuanto al término “grafo”, este fue introducido por James Joseph Sylvester²³ en un artículo publicado en la revista *Nature*: “Chemistry and Algebra: On an Application of the New Atomic Theory to the Graphical Representation of the Invariants and Covariants of Binary Quantics”. En él, Sylvester mostró una analogía entre “invariantes cuánticos” y “co-variantes” del álgebra y los diagramas moleculares. (Sylvester, 1978)^[175]

²³ James J. Sylvester (1.814 – 1.897) Matemático inglés. Emigró a EEUU donde fundó el “American Journal of Mathematics”.

Every invariant and covariant thus becomes expressible by a graph precisely identical with a Kekuléan diagram or chemicograph. I give a rule for the geometrical multiplication of graphs, i.e. for constructing a graph to the product of in- or co-variant whose separate graphs are given. (Sylvester, 1978:284)^[175]

Sin embargo, Foulds indicó que la primera utilización del término “grafo” procede de la frase “anotación gráfica” (“*graphic notation*”), introducida por primera vez por E. Frankland y adoptada por A. Crum Brown en 1884.

Each atom of a chemical structure is represented by a point of a graph and atomic bonds are represented by lines. (Foulds, 1992:XXIII)^[163]

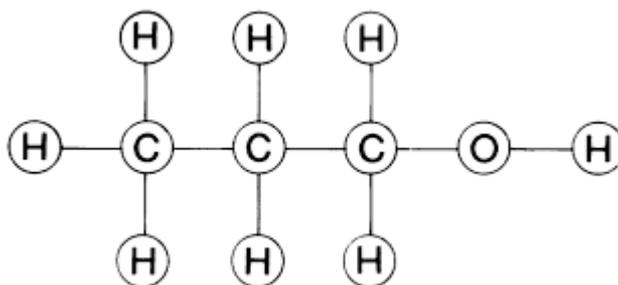


Figura 4: Isómeros químicos. (Foulds, 1992:XXIV)^[163]

1885:

En el transcurso del año 1885, Charles S. Peirce²⁴ denominaría “*existential graphs*” a un sistema lógico de notación a partir de un sistema iconográfico de símbolos fáciles de entender: “*entitative graphs*”. Por ejemplo, *and* (^) o *negation* (-). (Zhang, 2002:5)^[176]

El trabajo de Peirce ha tenido posteriormente una interesante aplicación planteada por John F. Sowa. Partiendo de la propuesta de Peirce, Sowa presentó el concepto “gráficos conceptuales” con varios usos en el campo del conocimiento informático. (Sowa, 2011)^[177]

1893:

Se considera que Émile Durkheim²⁵ y Ferdinand Tönnies fueron precursores en el avance hacia el concepto de las redes sociales tal y como las entendemos ahora. Tönnies argumentaba que dentro de los grupos sociales pueden existir dos tipos de lazos: personales y directos, que enlazan individuos que comparten valores y creencias (*Gemeinschaft*), o vínculos impersonales que son formales e instrumentales (*Gesellschaft*). Por su lado, Durkheim dio una explicación no individualista de la realidad social: coligió que los fenómenos dentro de una realidad social surgen cuando la interacción entre sujetos ya no puede explicarse en términos de las propiedades de los individuos por separado. En cuanto a la tipificación del tipo de sociedades, Durkheim distinguió entre una sociedad “tradicional” y una sociedad “moderna”. La sociedad tradicional se caracterizaría por presentar una “solidaridad mecánica” dentro de sí, que liga a individuos de características similares a través de lazos sociales normativos. En cambio, la sociedad moderna se fundamenta en la “solidaridad orgánica”, desarrollada a partir de la cooperación entre individuos diferenciados a través de la división del trabajo. (Freeman, 2012:14)^[178]

²⁴ Charles Sanders Peirce. (Massachusetts 1839 - Pensilvania 1914). Filósofo, lógico y científico estadounidense. Considerado el fundador del pragmatismo y de la semiótica moderna.

²⁵ David Émile Durkheim (1858-1917) Sociólogo, Psicólogo social y Filósofo francés, considerado como el padre de la sociología y precursor de las redes sociales.

Inicios del siglo XX:

Dentro del academicismo, Georg Simmel²⁶ sería también un pionero en términos de otorgar al concepto de red un papel decisivo para explicar los fenómenos sociales. Del filósofo alemán podemos destacar este enunciado que resume su relevante contribución al nexo entre “red” y “sociedad” desde el campo investigativo: “la sociedad existe ahí en donde un número de individuos interactúan” (Simmel, 2011:23)^[179]. En consecuencia, Simmel consideró que era menester de la sociología el estudio de sus patrones de interacción y funcionamiento.

1930:

La célebre teoría de los “Seis grados de separación”, de la que hablaremos posteriormente, tiene su origen en el año 1930 y la plantea Frigyes Karinthy²⁷ aunque sin referirse a ella con el nombre que le daría la fama varias décadas después. La importancia de Karinthy es significativa, puesto que su planteamiento coincidirá plenamente con el propuesto por Stanley Milgram que instaurará la denominación de “Teoría de los seis grados de separación” (Karinthy, 1929)^[180]. Ya en 2004, D. Watts expandirá la teoría e incorporará nuevos elementos acorde con los cambios en las interacciones sociales y las dinámicas relacionales que ha propiciado la reciente revolución tecnológica y sus ramificaciones en campos como la comunicación.

1930 - 1940:

Entre los años treinta y cuarenta del siglo pasado, los estudios llevados a cabo en Harvard a cargo de los investigadores L. Warner y E. Mayo contribuyeron de forma decisiva a fijar y explicar los orígenes de las redes sociales. Sus investigaciones son conocidas por haber fijado como campo de estudio las estructuras de trabajo de una fábrica eléctrica de Chicago. (Schramm, 2012)^[181]

²⁶ Georg Simmel. (Berlín 1858 - Estrasburgo 1918). Filósofo y sociólogo alemán.

²⁷ Frigyes Karinthy. (Budapest 1887 - Siófok 1938). Escritor húngaro.

En esta misma década, desarrolla su labor en el campo de las relaciones sociales el antropólogo social inglés Radcliffe-Brown²⁸. Su trabajo enfatizará el papel de las relaciones sociales y que él considera reales y directamente observables. Su apunte significativo es el que logró al separar esa realidad “tangible” de las relaciones sociales de lo que sería la “estructura social”, una construcción teórica postulada por los científicos y que se basa justamente en el estudio de la “observación” de las relaciones sociales en cuanto a sucesos reales. Radcliffe-Brown vio en las matemáticas discretas una herramienta apropiada y necesaria para la antropología social, la joven disciplina de estudio con el objeto de identificar las estructuras sociales y las relaciones entre ellas. En la misma línea que Simmel, Radcliffe-Brown defendió que la sociedad podía ser vista como una red compleja de relaciones sociales que bautizó “estructura social total” (Radcliffe-Brown, 1940:3)^[182]. Dicha estructura es la suma del total de las relaciones sociales en una unidad que las engloba, la cual podrá ser analizada y cuantificada con el uso de las matemáticas aplicadas al ARS. (Prell, 2012)^[17]

1936:

En 1936, Dénes König²⁹ publica en Alemania el primer libro sobre la Teoría de Grafos moderna: *Theorie der endlichen und unendlichen Graphen* (Grimaldi, 1998)^[183]. Por su lado, el psicoanalista Kurt Lewin³⁰ se daría a conocer durante la misma década de los treinta por haber desarrollado el marco teórico entendido como la “teoría de campo” y que describe y explica el comportamiento humano desde una perspectiva estructural. Por ejemplo, Lewin propondría que el espacio interior de un individuo es representable por un mapa plano. En la actualidad, sus trabajos están siendo interpretados por medio de grafos y ello ha permitido que más psicoanalistas utilicen los grafos como interpretación psicológica de un individuo y su ambiente situando las personas como nodos y sus relaciones personales como aristas o enlaces. (Prell, 2012:24)^[17]

²⁸ Alfred Reginald Radcliffe-Brown (1881- 1955). Antropólogo social inglés, desarrolló la Teoría del Funcionalismo estructural.

²⁹ Dénes König. (Budapest 1884 – 1944). Matemático húngaro.

³⁰ Kurt Lewin. (Polonia 1890 - Estados Unidos 1947). Psicólogo, médico y filósofo.

En otra contribución relevante de aquel tiempo, el antropólogo Ralph Linton introduciría la idea del “rol social”, en la que dos individuos, como por ejemplo padre e hijo, ocuparían, primero uno y en su madurez el otro, posiciones similares. Asimismo, será probable que aparezcan en ellos algunos comportamientos similares. (Turner, 2006)^[184]

1937:

Manteniéndonos todavía sobre los mismos años, Jacob L. Moreno³¹ seguiría el camino asentado en su momento por Durkheim y Tönnies y publicaría el año 1934, en colaboración con Helen Jennings, la obra que daría al término “red” el sentido que le atribuimos en nuestra contemporaneidad. (Freeman, 2012:35)^[178]

A su vez, Jacob L. Moreno fundaría la “Sociometría” (*Sociometry*) para la medición de las relaciones interpersonales en grupos pequeños e inventaría el concepto *sociogram* para el estudio de las relaciones interpersonales como estructuras, en las cuales, las personas o cualquier unidad social están representadas por puntos en un espacio bidimensional; las relaciones entre dichas unidades sociales se representan por líneas que enlazan los puntos correspondientes. (Prell, 2012)^[17]

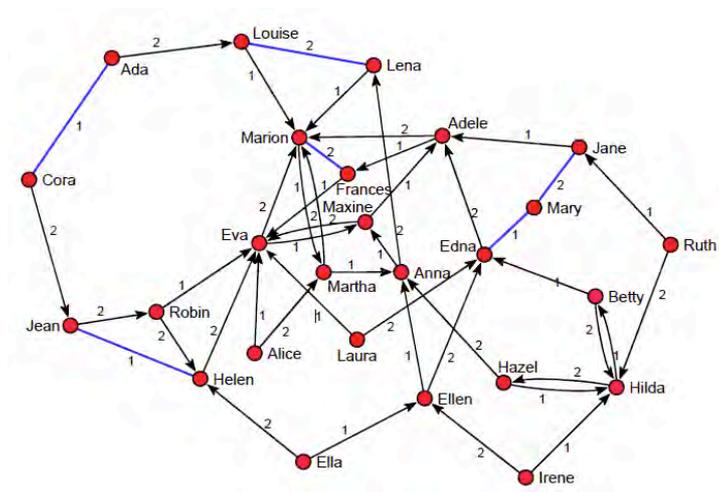


Figura 5: Sociograma que representa un estudio sociométrico de la estructura de un grupo. En este caso, de las preferencias en la elección de una compañera de mesa. (De Nooy, 2011:4)^[185]

³¹ Jacob L. Moreno (1.889 – 1.974) Nació en Rumanía y desarrolló su profesión EEUU. Psiquiatra y fundador de la sociometría.

1940:

Max Gluckman³² de la escuela antropológica de Manchester, seguidor del estructural-funcionalismo y considerado su figura más destacada, desarrolló el principio de los “lazos transversales” (*cross-cutting ties*), relaciones “transversales” o alianzas, que se basa en que los conflictos son inevitables en la sociedad e incluso sirven en el mantenimiento de la misma. Sostenía que los grupos que forman una sociedad tienen la tendencia a romperse para posteriormente volverse a unir mediante nuevas alianzas. Las sociedades se rompen, se reforman pero el sistema social, en su conjunto, se mantiene. (Prell, 2012)^[17]

1949:

A. Rapaport³³. Desarrolló un enfoque relacional y un acercamiento probabilístico para la caracterización de grandes redes sociales, en las que los nodos son las personas y los vínculos son de amistad o conocimiento. A partir de parámetros de conectividad cercana, desarrolló el siguiente razonamiento: si A está enlazado a B y C, hay una gran probabilidad que B y C estén o acaben estando vinculados entre sí, a partir de la gran regla de conectividad global de la red. (Solomonoff, 1951:107)^[186]

1950:

Alexander Bavelas³⁴, como director del *Group Networks Laboratory* del *Massachusetts Institute of Technology (MIT)* desarrolló las primeras investigaciones sobre la “centralidad”, la cual la relacionaba con la eficiencia de un grupo a la hora de resolver problemas, situaciones de liderazgo o la cohesión y el grado de satisfacción de sus miembros. (Freeman, 2000:6)^[187]

³² Max H. Gluckman (1911–1975) Antropólogo británico, nacido en Sudáfrica. Realizó grandes contribuciones en el análisis de la antropología política y los sistemas políticos entre las tribus africanas.

³³ Anatol Rapaport (1.911 – 2.007) Ruso de nacimiento, nacionalizado USA. Psicólogo matemático. Contribuyó en: biología matemática, interacción social, contagio, análisis de las redes sociales y Game Theory.

³⁴ Alexander Bavelas (EEUU 1.920) Psicosociólogo Americano. Fundador del Group Networks Laboratory del MIT.

Ithiel de Sola Pool³⁵ y Manfred Kochen, escribieron un manuscrito: “Contactos e Influencias” (finalmente publicado en 1978), en el que se planteaban diversas cuestiones sobre redes sociales, una de las cuales era: “what is the probability that two strangers will have a mutual friend”, o el número de grados de separación en las actuales redes sociales. Testigo que aceptó Milgram para resolver esta cuestión. (Ver S. Milgram) (Pool, 1978:14)^[188]

1950 - 1960:

El Análisis de las redes sociales (ARS) se desarrolló con los estudios de parentesco de Elizabeth Bott en la Inglaterra de los años 1950 y con los estudios de urbanización del grupo de antropólogos de la Universidad de Manchester (acompañando a Max Gluckman y después a Clyde Mitchell) entre los años 1950 y 1960, investigando redes comunitarias en el sur de África, India y el Reino Unido.

Al mismo tiempo, el antropólogo británico Frederick Nadel, codificó una teoría de la estructura social, que influyó posteriormente en el análisis de redes. (Nadel, 1957)^[189]

1954:

J. A. Barnes es el antropólogo a quien se le atribuye el concepto: “Red Social” (Social Network) que utilizó de forma sistemática para establecer patrones de “aristas” uniendo conceptos utilizados por el público con los utilizados por los científicos, como: segmentos (tribus, familias) y categorías sociales (género o etnia). Barnes analizó la importancia de las relaciones informales como la amistad, el parentesco, la vecindad, siendo la vida social total, como una suma de redes parciales de relaciones interpersonales. (Freeman, 2004:10)^[190]

³⁵ Ithiel de Sola Pool (1917 – 1984) Revolucionario en el campo de las ciencias sociales. Creador del término “convergencia” relacionada con la tecnología y comunicación.

1956:

Noam Chomsky³⁶ fue uno de los primeros investigadores del análisis sintáctico del lenguaje natural (*Natural Language Parsing*), el padre de la lingüística moderna y el introductor de la idea de que las lenguas pueden ser interpretadas como conjuntos de enlaces. Sus estudios en las reglas de escritura se conocen como la “Gramática de Chomsky” y los diversos trabajos que ha realizado tienen una gran influencia en las aplicaciones de resúmenes de textos. (Manning, 2000)^[191]

1959:

Paul Erdős³⁷ & Alfred Rényi proponen la *Theory of random graphs*, una combinación entre la Teoría de Grafos y la Teoría de la probabilidad que se hará imprescindible para el modelado de redes complejas. (Bollobás, 1998)^[192]

1960:

Durante el período en el que Harrison White³⁸ dirigió el Departamento de Sociología de la Universidad de Harvard este se convirtió en un centro de análisis de redes sociales de gran importancia e influencia futura. Las investigaciones de White y su equipo aportaron un nuevo enfoque al modo de entender las redes sociales como redes completas y desarrollaron nuevas técnicas matemáticas para descubrir estructuras dentro de las relaciones sociales. (Prell, 2012:44)^[17]

³⁶ Noam Chomsky (EEUU 1.928) Profesor emérito de lingüística del MIT. Estudios en teoría lingüística y ciencia cognitiva.

³⁷ Paul Erdős (1.913 – 1.996) Matemático Húngaro. Considerado el matemático más prolífico del s. XX.

³⁸ Harrison C White (EEUU, 1.930) Profesor de Sociología. De gran influencia en el desarrollo de teorías sobre las redes sociales y modelos matemáticos de las estructuras sociales.

1965:

Será Murray Beauchamp el primer matemático en definir el concepto de “cercanía”, una de las medidas radiales de longitud más conocidas y utilizadas calculada a partir del promedio de las distancias más cortas desde un nodo hacia todos los demás. Esta medida permite definir el nodo más “cercano” al centro de la red. (Beauchamp, 1965)^[193]

1967:

Stanley Milgram³⁹ desarrolla un experimento bajo el título “El Problema del Mundo Pequeño” (*The Small World Problem*) y con el que demuestra que la sociedad humana es una red social con la estructura de un mundo pequeño en el que las interconexiones son mucho más cortas de que lo que se piensa y ratifica, con ello, una célebre expresión popular: ¡Qué pequeño es el mundo!

El resultado de su experimento demuestra que cualquier persona puede alcanzar a cualquier otra del mundo en un promedio de 5 a 7 intermediarios. Los descubrimientos de Milgram fueron publicados en *Psychology Today* e inspiraron la expresión “Seis grados de separación”. (Prell, 2012:46)^[17]

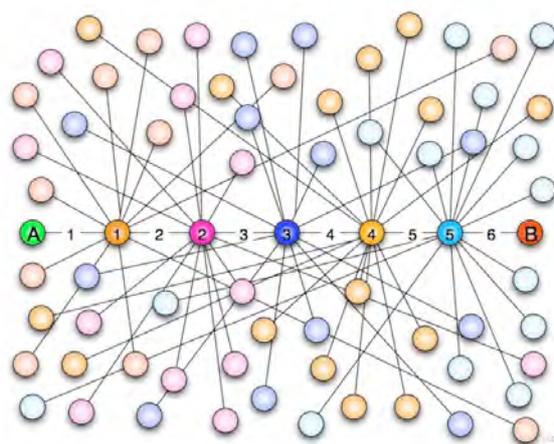


Figura 6: Ilustración de los grados (personas) que separan a la persona A de la B, desconocidos entre sí. (Wikipedia, 2014)^[194]

³⁹ Stanley Milgram (EEUU, 1.933 – 1.984). Se considera uno de los psicólogos más importantes del s. XX. Autor del experimento: “*Small Worlds*”.

1969:

J. C. Mitchell⁴⁰ pone las bases sistemáticas para el análisis de las redes sociales como convergencia de las escuelas estructural-funcionalista antropológica (representadas, entre otros, por Gluckman y Barnes) y la proveniente de la Teoría de Grafos (representada por Frank Harary⁴¹ y Bavelas, entre otros). (Wasserman, 2009)^[195]

El mismo año, Frank Harary publica el libro *Teoría de Grafos*. Entre muchas de sus aportaciones, Harary desarrolla la teoría de los grafos etiquetados (*Signed Graphs*), en la que los conceptos se representan como nodos etiquetados y conectados por aristas (arcos). Ello crea lo que el propio autor denomina “forma del grafo total” (*Total Graph Form*). El autor aplica la Teoría de Grafos a problemas con una estructura social. (Harary, 1969)^[196]

A su vez, Heindrich Heesch publica un método para resolver el problema de los cuatro colores utilizando ordenadores. El problema había permanecido sin resolver desde que lo formuló Francis Guthrie en el año 1852. (Fritsch, 1998)^[169]

1970:

En la década de los 70 se continúa profundizando en la labor de investigación en el campo del análisis de las redes sociales iniciado por White en el MIT la década anterior. Asimismo prosigue el desarrollo de la “Teoría de los seis grados de separación”. Su aportación, durante este período, representará un cambio fundamental en la metodología y será conocido como el “estructuralismo de Harvard”. Dicho estructuralismo desarrolla los análisis de las redes sociales como un método de análisis estructural que parte de modelos algebraicos, la Teoría de Grafos y otras técnicas como la escala multidimensional. (Freeman, 2012)^[178] (Prell, 2012)^[17] (Bergs, 2005)^[197]

⁴⁰ James Clyde Mitchell (1918 – 1995) Sociólogo y antropólogo británico. Investigador en Redes Sociales en la Universidad de Manchester.

⁴¹ Frank Harary (EEUU 1.921 – 2.005) Matemático y considerado como uno de los padres de la Teoría de Grafos moderna.

Otra de las tendencias en auge será la “Red de transición aumentada” (*Augmented Transition Network, ATN*), creada por W. A. Woods para el análisis del lenguaje natural. Tendrá importantes repercusiones en la lingüística computacional y en posteriores modelos de procesamiento de las oraciones humanas. (Woods, 1970)^[198]

1972:

Philip Bonacich será el primero en definir la “centralidad de vector propio” (*Eigenvector Centrality*) (ver Aplicaciones de la Teoría de Grafos) (Bonacich, 1987)^[199]

1973:

Mark Granovetter⁴² presenta la teoría de “La Fuerza de los lazos débiles” (*The strength of weak ties*) que aparece publicada en *The American Journal of Sociology*. Este trabajo, surgido a partir de la tesis doctoral de Granovetter, se basa en datos obtenidos acerca de cómo los profesionales de Massachusetts utilizaron sus redes personales para encontrar información sobre las nuevas oportunidades de trabajo. Según Granovetter, un lazo débil es aquel con el que no se tiene un vínculo emocional y con el que se mantiene poco o ningún contacto. Son la clara diferenciación de los lazos fuertes, como la familia y los amigos. El estudio demuestra que los profesionales consiguieron más trabajos por la vía de los lazos débiles que por la de los lazos fuertes. (Prell, 2012:76)^[17]

1979:

Linton Freeman⁴³ es el primero en dibujar un gráfico que resume, en términos visuales, los distintos significados de la “centralidad”. Con ello, popularizará el concepto de cercanía. Será el creador de una figura llamada grafo de la estrella (*star graph*). (Prell, 2012:96)^[17]

⁴² Mark Granovetter (EEUU, 1943). Sociólogo y profesor de la Universidad de Stanford. Elaboró la teoría de “La fuerza de los lazos débiles”.

⁴³ Linton C. Freeman (EEUU). Sociólogo y Profesor de la Universidad de California.

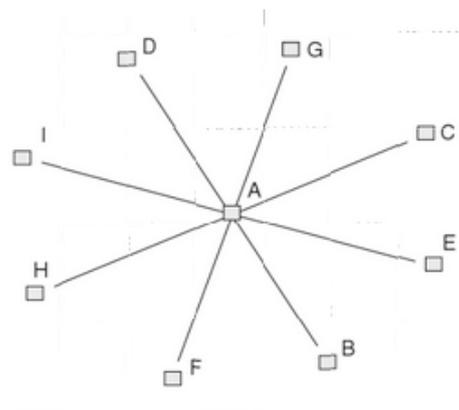


Figura 7: Ejemplo de “Grafo Estrella” en el que la persona A es la “estrella” que atrae la atención del resto del grupo. (Prell, 2012:96)^[17]

1980:

A principios de los años 80 se produce un importante desarrollo en la teoría sobre los análisis gramaticales. Teorías como la Gramática Léxico Funcional (*Lexical Functional Grammar, LFG*) y la Gramática Unificada Funcional (*Functional Unification Grammar, FUG*) contribuyen al desarrollo del papel de la información léxica en el análisis automático de textos. Por ejemplo, el Analizador Experto de Word “*Word Expert Parser*” (*WEP*) ha sido desarrollado prestando especial atención a los posibles diferentes significados de las palabras y analiza fragmentos de textos de lenguaje natural. Por otra parte, en esta década hay un resurgimiento de los enfoques estadísticos o empíricos para el procesamiento del lenguaje natural, de entre los que destacan los avances en reconocimiento de voz (Rabiner, 1989)^[200], la señalización de las partes de un discurso a través de la identificación de palabras clave (Charniak *et al.*, 1993)^[201], el análisis sintáctico (Ratnaparkhi, 1999)^[202] o la segmentación de discursos y textos. (Litman, 1996)^[203] (Zhang, 2002:10)^[176]

1982:

Aparece la teoría de la Ciencia de los Grafos (*Knowledge Graphs*) a cargo de Hoede y Stokman, quienes extraen conocimientos sobre textos médicos y sociológicos para desarrollar sistemas expertos. Al inicio, solo diferenciaron tres tipos de relaciones, siendo la relación causal la más importante. (Zhang, 2002)^[176]

1984:

John F. Sowa⁴⁴ crea los Gráficos conceptuales, representación gráfica de las redes semánticas de la inteligencia artificial y los gráficos existenciales de Charles S Peirce. (Roberts, 1973)^[204], (Sowa, 2011)^[177]

1990:

Se registra un aumento del interés en el desarrollo de modelos estadísticos para el análisis de datos de redes sociales. El modelo estadístico más famoso será la familia de los modelos conocida como modelos aleatorios de gráficos exponenciales (*Exponential Random graph models, ERGMs*), creados a partir del modelo de Markov. (Prell, 2012)^[17] (Robins, 2006)^[205]

1991:

John Scott⁴⁵ desarrolla una gran labor de investigación de las redes sociales y su origen histórico, contribuyendo a popularizar el método del análisis estructural a través de la síntesis de textos. (Scott, 1991)^[166]

1992:

Ron Burt⁴⁶ desarrolla el concepto de Agujeros estructurales (*Structural holes*) para discutir cómo algunas estructuras de redes pueden generar una ventaja estratégica a algunos individuos respecto a los demás. Burt se refiere a los agujeros estructurales como a los espacios vacíos en las estructuras sociales. Los espacios vacíos son fruto de individuos (nodos) que no tienen un enlace (arista) entre sí. (Burt, 2004)^[206]

⁴⁴ John F. Sowa (EEUU, 1940) Matemático e informático estadounidense. Inventor de los gráficos conceptuales. Investigador en inteligencia artificial y automatización del lenguaje natural.

⁴⁵ John Scott (London 1949) Sociólogo inglés.

⁴⁶ Ronald S. Burt (1949) Sociólogo americano. Investigador en redes sociales y capital social.

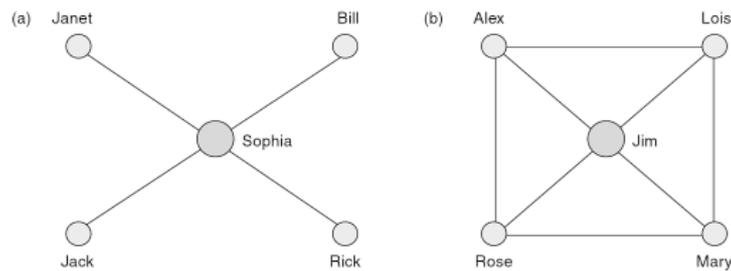


Figura 8: ilustración de dos posibles situaciones de agujeros estructurales, de tal modo que Sofía, en su red personal, hace de puente de cuatro agujeros estructurales y Jim, en cambio, no hace de puente de ningún agujero estructural. Como afirmará Burt, Sofía puede tomar mejores decisiones que Jim dentro de su ámbito laboral al tener acceso a mayor y mejor información. (Prell, 2012:123)^[17]

1993:

S. Wasserman trabaja en pequeños grupos a partir de los grafos teóricos para analizar su estructura social pone de manifiesto de qué manera la estructura del grupo afecta al comportamiento individual. (Wasserman, 1994)^[207]

1998:

Duncan Watts⁴⁷ y Steven Strogatz publican un artículo en la revista *Nature* denominado *Collective dynamics of Small-World networks*. Del artículo derivará el “Modelo Watts & Strogatz” para la construcción de redes en la Teoría de Grafos. (Watts, 1998)^[208]

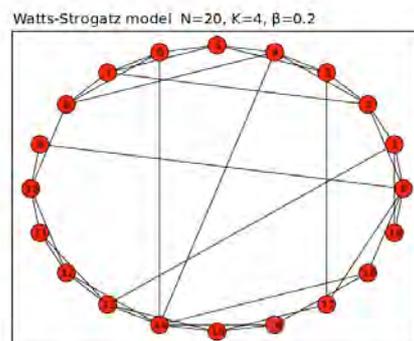


Figura 9: Red de 20 nodos construida según el modelo de Watts & Strogatz. (Wikipedia, 2014)^[209]

⁴⁷ Duncan J. Watts (Canadá 1971) Profesor de Sociología de la Universidad de Columbia. Autor del libro “Seis Grados” y del concepto “Small-Worlds”, con grandes influencias en el análisis de las redes sociales.

Su aportación será revolucionaria y significará la entrada en el estudio de las Redes Sociales de investigadores del ámbito de la física. Ello permitirá cuantificar los *Small Worlds* usando métodos estadísticos para el análisis de redes sociales.

En 1999 Albert-László Barabási y Réka Albert realizaron una importante contribución en la investigación de la centralidad de grado. (Barabási, 2002)^[210]

2000:

Scott organiza el desarrollo del Análisis de Redes Sociales a partir de los avances tecnológicos, como por ejemplo, el desarrollo de la utilización de la Teoría de Grafos y la investigación sobre pequeños y exclusivos grupos de personas. (Prell, 2012:36)^[17]

2001:

Robert Putnam populariza el análisis de las redes sociales a raíz de la publicación de sus libros sobre el capital social y el papel de las redes sociales en la construcción de una comunidad. Su principal aportación y repercusión por parte de académicos y políticos se ha centrado en la distinción entre *bridging* y *bonding*, siendo *bridging* (puente) las estructuras del Capital Social formadas por redes particulares forjadas por lazos débiles y que mantienen estructuras abiertas. Por su lado, las redes llamadas *bonding* (unión) están constituidas por lazos fuertes y por estructuras de redes compactas. (Prell, 2012:46)^[17]

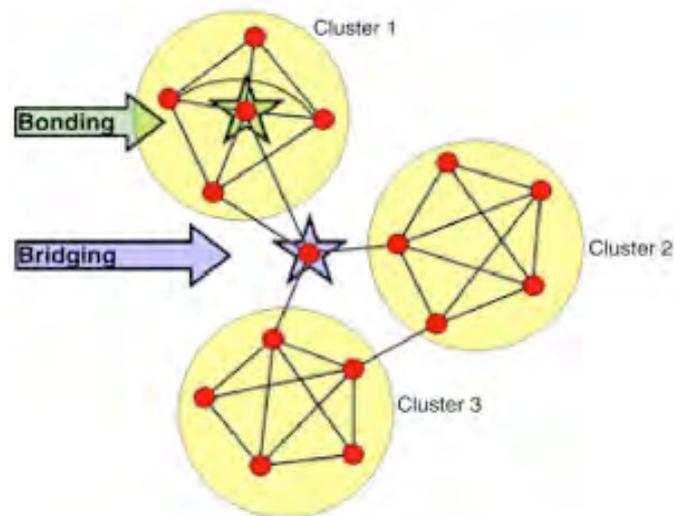


Figura 10: *Bonding* (también denominado *closure*) y *bridging* (también denominado *brokerage*) son dos formas diferentes de conectividad. *Bonding* corresponde a conexiones dentro de un grupo muy unido, mientras que *bridging* corresponde a conexiones con otros grupos. *Bonding* denota un sentido de comunidad sólida en el que las interacciones son comunes y eficientes, mientras que *bridging* indica el acceso a nuevos recursos y oportunidades. (Hoppe, 2010:602)^[211]

2003:

Moody y White utilizan la Teoría de Grafos para definir la *Cohesión estructural*, que definen de la siguiente manera: “número mínimo de actores que, si se eliminan de un grupo, harán que este se desconecte o desaparezca”. Presentaron la definición del “arraigo” o “integración social” en términos de una agrupación jerárquica de cohesión de las estructuras, renovando y mejorando la estructura tradicional del modelo de Red Social. (Moody, 2003:3)^[212]

2004:

Duncan J. Watts publica el libro *Seis Grados: la ciencia de una edad conectada* (*Six Degrees: the science of a connected age*). En su interior quiere probar que, en la Tierra, cualquier persona del planeta puede estar conectada con cualquier otra a través de una cadena de no más de 5 personas conocidas y conectadas entre sí con sólo seis enlaces. (Watts, 2004)^[44]

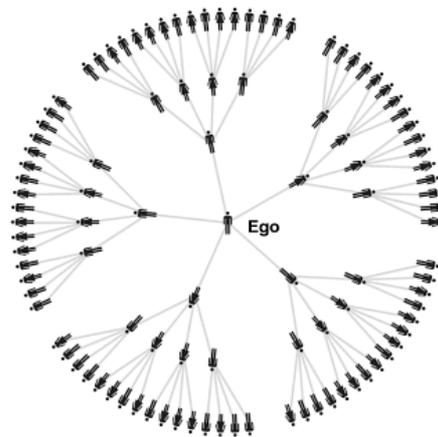


Figura 11: Gráfico que ilustra la hipótesis que formuló Milgram y que se desarrolla en el libro de Watts. En él, la figura “Ego” tiene cinco amigos y cada uno de ellos tiene cuatro amigos más. A su vez, el segundo grupo de cuatro tiene, cada uno de sus integrantes, otros cuatro amigos. “Ego” puede alcanzar en dos grados a 25 personas. Si “Ego” tuviese 100 amigos y cada uno de ellos tuviese a su vez otros 100 amigos, con tan solo tres grados de separación “Ego” podría alcanzar a casi 1.000.000 de “amigos”. Así sucesivamente hasta alcanzar a la casi totalidad del planeta en los seis grados de separación. (Watts, 2004:40)^[44]

2006:

Mark Newman, destacado investigador en redes sociales y de la información, introduce junto a Girvan el concepto de “Modularidad”. Este concepto permite establecer un índice del grado en que cada nueva partición forma un grupo o alguna otra forma de comunidad. Presentan el Método Girvan-Newman que explica la sucesiva eliminación de enlaces de alta intermediación. Los enlaces de alta intermediación son los que, en todos los pares de nodos, llevan el mayor volumen de tráfico a lo largo de los caminos más cortos. Para ello se basan en la premisa que éstos son los enlaces más “vitales” para conectar las diferentes regiones de una red. (Jin, 2001)^[213]

2009:

Sebastian Schnettler, sociólogo alemán de la Universidad de Konstanz, investiga el hecho de que personas que no se conocen puedan tener amigos en común y con ello logra demostrar que los grados de separación o distancias de los trayectos (enlaces entre personas) pueden ser modelados en una red social. (Schnettler, 2009)^[214]

2011:

Lars Backstrom, de la empresa Facebook, realiza un estudio denominado *Four Degrees of Separation* con todos los usuarios activos de la empresa página en ese momento: 721.000.000 miembros (alrededor del 10% de la población mundial). En el estudio, Backstrom analiza el conjunto de amigos en común, saca el promedio de cuántos eslabones hay entre cualquier usuario y otro cualquiera, excluyendo de este promedio a celebridades y famosos. Los resultados muestran que el 99,6% de pares de usuarios están conectados por entre 4,4 y 5,7 eslabones, en función de la muestra elegida. (Backstrom, 2011)^[215]

3.2. Definición y componentes de un grafo

3.2.1. Definición

La Teoría de Grafos estudia un tipo de estructuras matemáticas (grafos) que se utilizan para establecer relaciones entre pares de objetos de un mismo conjunto o segmento.

Un “grafo” está formado por un conjunto de “nodos” (*nodes*, “puntos” o “vértices”), representados por una serie de puntos y un conjunto de “aristas” (*edges* o “enlaces”) que conectan pares de nodos mediante líneas trazadas. En esta representación, un grafo se define como un par ordenado $G = (V, A)$ que comprende un conjunto V de vértices o nodos y un conjunto A de aristas o enlaces. (Singh, 2010:116)^[216] (Biggs, 1986:9)^[217] (Chartrand, 1977:253)^[173]

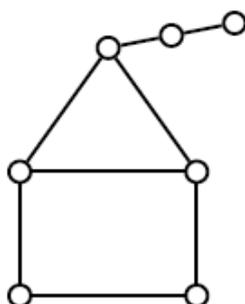


Figura 12: representación de un grafo conocido con el nombre de “House”.

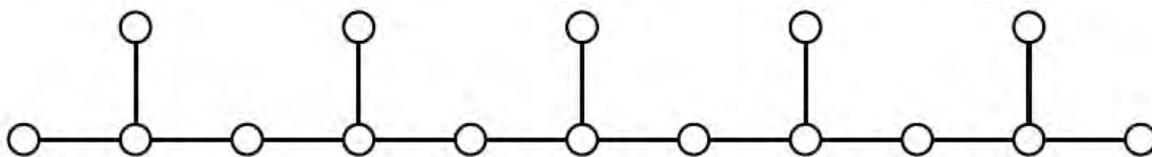


Figura 13: representación de un grafo

3.2.2. Componentes de un grafo

Orden de un grafo:

Un grafo G es un par ordenado $G = (V, A)$, donde:

- $V =$ es un conjunto de vértices o nodos, y
- $A =$ es un conjunto de aristas o arcos que relacionan dichos nodos.

Normalmente, V suele ser finito. Se llama orden del grafo G a su número de vértices (nodos), designado como “ V ”. El grado de un vértice o nodo V es igual al número de arcos A que se encuentran en él. (Chartrand, 1977:574)^[173]

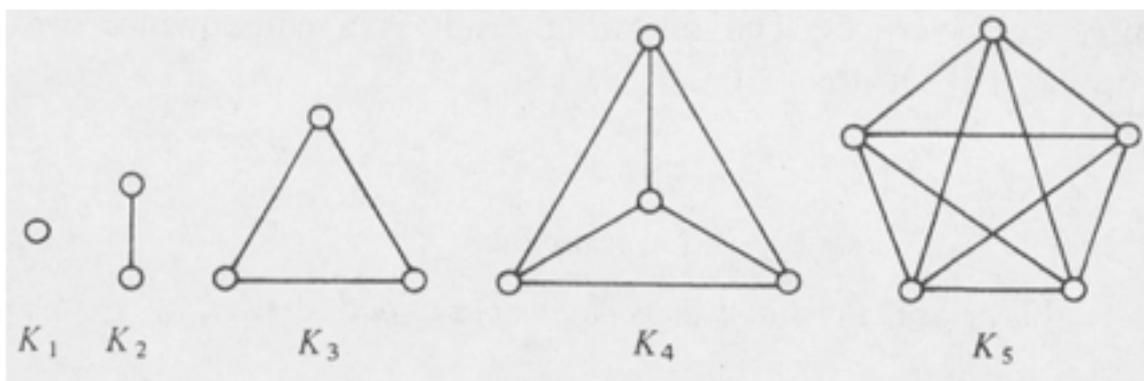


Figura 14: Ejemplos de grafos

Grafo Dirigido o Dígrafo (*Digraph*):

Un grafo dirigido o dígrafo es un grafo en el que las aristas (*edges*) están dirigidas. Formalmente, un dígrafo es (normalmente) un número finito de nodos (Vértices: V) y sus pares (aristas) ordenados (a, b) (donde a y b forman parte de V). Cada arista dirigida es un enlace que va de un nodo a otro, siendo importante hacia qué dirección se dirige. (Bang-Jensen, 2008:3)^[218]

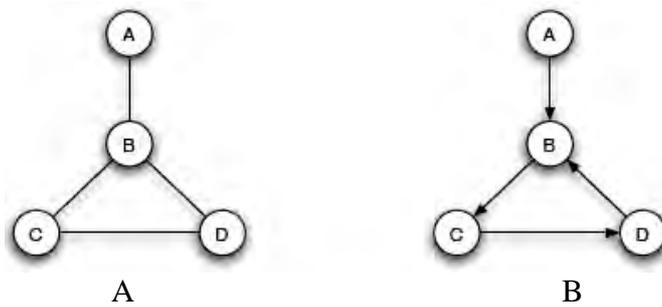


Figura 15: Ejemplo de dos grafos: (A) no dirigido y (B) dirigido. (Easley, 2010:24)^[219]

Paseos (Walks):

Secuencia de nodos con la particularidad de que cada par de nodos está conectado por una arista. Dentro del concepto “Paseo” se desarrollan diferentes variaciones:

- *Path* = camino. Tipo de paseo en el que no se pasa dos veces por el mismo nodo, a excepción de que coincida con el inicio y el final del trazado.
- *Cycle* = Ciclo. Tipo de paseo que inicia y finaliza en el mismo nodo.
- *Trail* = Pista. Tipo de paseo que no repite ninguna vez la misma, exceptuando si su procedencia es de diferentes nodos.
- *Circuit* = Circuito. Tipo de paseo *trail* con la diferencia de que empieza y finaliza en el mismo nodo.

El concepto de “camino” parte de la idea de que los elementos viajan a través de las aristas de los grafos, desplazándose de nodo en nodo en una secuencia (como la secuencia de enlaces aéreos que puede realizar un pasajero en un viaje, la información que puede realizar cierta información que viaja de persona a persona a través de una red social o el recorrido de un ordenador o un programa de *software* que visita diferentes sitios web siguiendo los enlaces). (Easley, 2010:26)^[219]

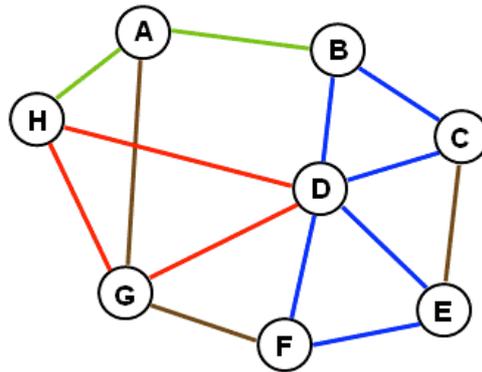


Figura 16: Grafo que representa varios tipos de paseos, señaladas por los distintos colores. (Wikipedia, 2007)^[220]

Distancia (*Distance* o *Length*):

La longitud de un camino es el número de nodos que atraviesa a lo largo de su recorrido.

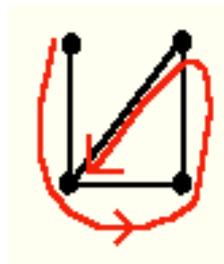


Figura 17: la longitud del camino de esta ilustración es = 4. (Caldwell, 1995)^[221]

A cada camino se le asocia una *distancia*, igual al número de aristas de ese camino. Se denomina *geodésica* al camino más corto entre un par de puntos. Se dice que los puntos *median* cuando se encuentran en la única geodésica del recorrido o en todas las geodésicas que unen un par de puntos. Por lo tanto, la distancia de un camino será la cantidad de pasos entre el principio y el final, y ello corresponde al número de vértices de la secuencia que lo comprende.

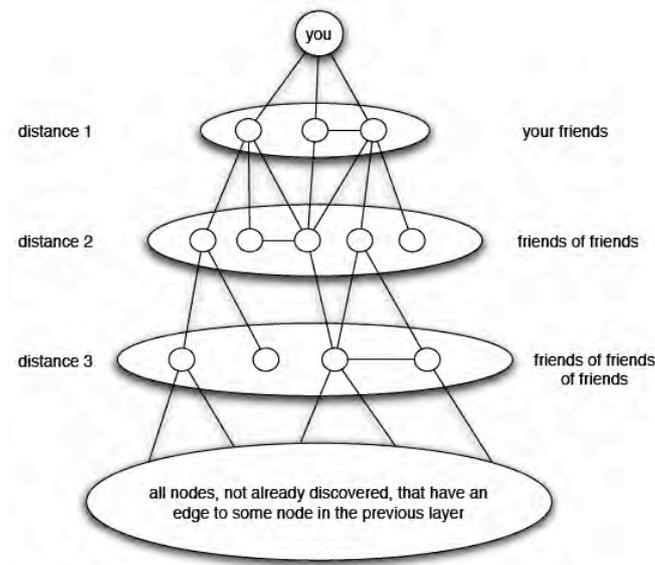


Figura 18: Distancias establecidas por “capas”. Cada capa está construida por nodos que tienen como mínimo un borde de al menos un nodo en la capa anterior. (Easley, 2010:33)^[219]

Conectividad (*Connectivity*):

Un grafo está *conectado* si existe un camino entre cada par de nodos. (Bollobás, 1978:2)^[222]

Árbol grafo (*Tree*):

El árbol grafo define a un grafo sencillo no dirigido G que no forma un círculo en el que dos vértices estén conectados por exactamente un único camino.

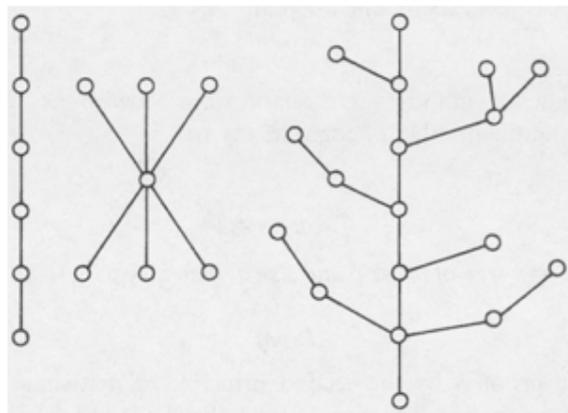


Figura 19: Ilustración de un grafo denominado “Árbol”. (Chartrand, 1977:1422)^[173]

Centralidad (*Centrality*):

En la Teoría de Grafos y en el Análisis de Redes Sociales (ARS) existen varias maneras para medir la centralidad de un vértice dentro de un grafo y su importancia relativa dentro del mismo. Por ejemplo, el estudio de la centralidad sirve para determinar la importancia que tiene una persona dentro de una red social o una determinada carretera dentro de una red de transporte.

Existen cuatro criterios ampliamente utilizados para medir la centralidad en el ARS: (Prell, 2012:103-104)^[17]

- Centralidad de Grado (*Degree Centrality*)
- Cercanía (*Closeness*)
- Intermediación (*Betweenness*)
- Centralidad de vector propio (*Eigenvector centrality*)

Centralidad de Grado (*Degree Centrality*):

Corresponde a un análisis cuantitativo y de volumen de medición que suma el número de enlaces (aristas) o contactos inmediatos que un actor (nodo) tiene en una red.

Para medir la centralidad de un grado, se cuenta el número de enlaces adyacentes al nodo, haciendo caso omiso de la dirección y el valor del enlace.

En el análisis de una red de comunicación, un actor (nodo) con alta centralidad de grado es aquel que puede ser considerado importante, un *influencer* que logra que la información fluya en la red y que tiene una gran capacidad para la con rapidez. (Prell, 2012:97)^[17]

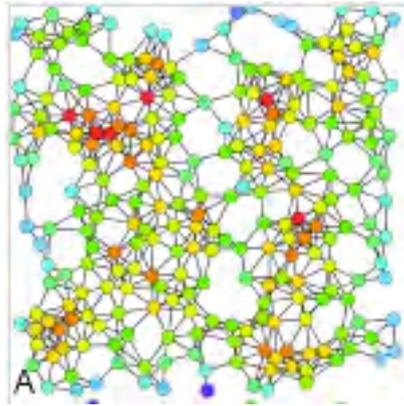


Figura 20: Ilustración de la centralidad de grado. El color rojo representa el máximo valor de centralidad y el azul el mínimo. (Wikipedia, 2014)^[223]

Intermediación (*Betweenness*):

La intermediación mide el lugar que ocupa un nodo entre otros nodos en una red. Esta medida tiene en cuenta la conectividad de los vecinos del nodo y otorga un mayor valor a los nodos que unen diferentes grupos. La medida refleja el número de personas a las que una persona puede conectarse de forma indirecta a través de sus enlaces directos. (Brandes, 2008:137)^[224]

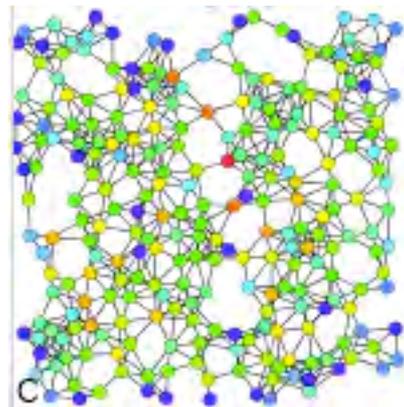


Figura 21: Ilustración de la Intermediación o *Betweenness*. El color rojo representa el máximo valor de intermediación y el azul el mínimo. (Wikipedia, 2014)^[225]

Cercanía (*Closeness*):

Se calcula con la suma o el promedio de las distancias más cortas desde un nodo hacia todos los demás. En algunos aspectos, la medida de cercanía difiere de las otras formas de centralidad. Mientras que la centralidad de grado destaca la actividad y la de intermediación el control del flujo de información, la cercanía se centra en la independencia del actor. La medida se puede describir a partir de la siguiente lógica: si un actor (nodo) no es central necesitará a otros para transmitir mensajes a través de la red (Bavelas, 1950)^[226] (Freeman, 1977:35)^[227]. Eso nos dice que un actor que esté cerca de muchos otros actores es un actor muy independiente. Este actor puede llegar rápidamente a los demás sin tener que depender tanto de los intermediarios.

Como se desprende de la medida cercanía, a un actor que tenga un alto grado de cercanía se le puede atribuir una mayor capacidad para movilizar más fácilmente una red al serle más sencillo llegar a los integrantes que la forman.

Además de la independencia, los investigadores han relacionado la cercanía con la capacidad de un actor para acceder fácilmente a la información de la red gracias a su poder e influencia. (Prell, 2012:107)^[17]

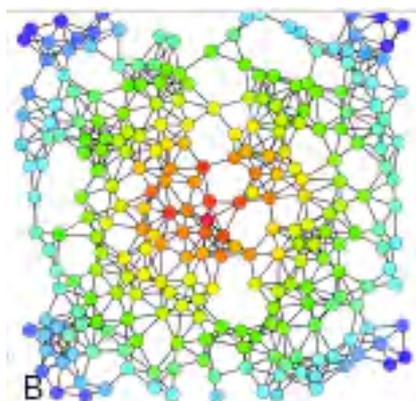


Figura 22: Ilustración de un ejemplo de centralidad cercanía o closeness. El rojo representa el máximo nivel de “cercanía” y el azul el mínimo. (Wikipedia, 2014)^[228]

Centralidad de Vector Propio (*Eigenvector Centrality*):

La Centralidad de Vector Propio mide la importancia e influencia de un nodo en una red. Su cálculo se obtiene asignando puntuaciones relativas o ponderadas a todos los nodos de la red, partiendo del principio de que los nodos con una puntuación más alta tienen mayores facilidades de conexión. Se establece aquí una diferencia entre la Centralidad de grado, en la que todos los nodos “pesan” igual y la Centralidad de Vector Propio, en la que los nodos “pesan” de manera diferente (Ruhnau, 2000:360)^[229]. Por ejemplo, el cálculo del *PageRank* del buscador Google utiliza éste criterio para medir la popularidad de los sitios web en Internet y mostrarlos (peso) según su importancia en los resultados de búsqueda. (Haveliwala, 2003:5)^[230]

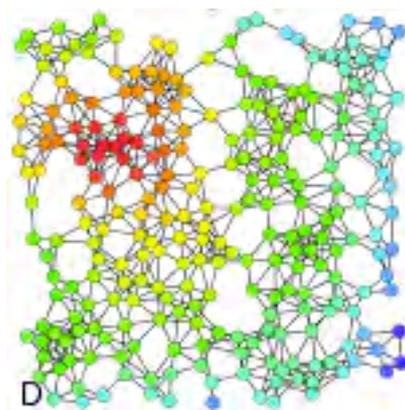


Figura 23: Ilustración de un ejemplo de Centralidad de Vector Propio. El rojo representa el máximo nivel de Centralidad y el azul el mínimo. (Wikipedia, 2014)^[231]

Matriz de adyacencias de un grafo:

Los grafos pueden representarse con frecuencia por su matriz de adyacencia:

ea $G = (N, A)$ un grafo de n nodos. La matriz de adyacencia M para G es una matriz $M_{n \times n}$ de valores booleanos, donde $M(i,j)$ es verdad si y solo si existe un arco desde el nodo i al nodo j .

Cuando el grafo no es dirigido, la matriz de adyacencia es simétrica. La matriz de adyacencia será la misma matriz de la relación A de N en N ya que indica qué nodos están relacionados (unidos por un arco).

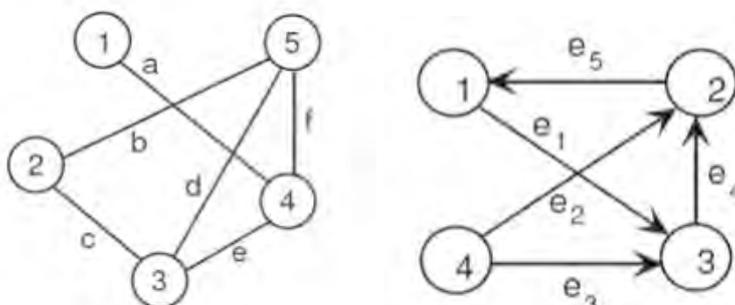


Figura 24: Representación de dos grafos y a sus correspondientes matrices de adyacencia. El ejemplo de la izquierda corresponde a un grado no dirigido y el de la derecha a un grado dirigido.

Sus matrices de adyacencias son: (Van Dooren, 2009:28)^[164]

$$A_1 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

Figura 25: Representaciones numéricas de matrices de adyacencias

Las filas y las columnas de la matriz representan los nodos del grafo. Como vemos en la matriz de adyacencia A1, la primera fila corresponde al nodo número 1 del primer gráfico y está conectado con el nodo número 4. Será así sucesivamente.

3.3. Aplicaciones de la Teoría de Grafos

El recorrido por los antecedentes históricos de la Teoría de Grafos nos ha permitido ir apuntando algunos de los usos a los que se ha ido aplicando, en especial cuando se ha aplicado a problemas y teoremas concretos, como la Teoría de los cuatro colores o la Teoría de los seis grados de separación. A medida que la Teoría de Grafos se desarrolla y estudia en diversos campos, cada vez de manera más sofisticada y nos adentramos en el nuevo milenio, observamos con interés su gran utilidad en el campo de las relaciones sociales. Para ello, ha sido crucial entender que una sociedad no es una suma de individuos y sus características, sino una estructura de relaciones interpersonales. Por ende, el individuo no es la unidad central básica: la estructura social se define tanto por el individuo como por sus relaciones. Las unidades centrales forman parte de grupos que, con sus interacciones, dan lugar a lo que entendemos por sociedad. (Nadel, 1957:12)^[189] (Otte, 2002:442)^[232]

Siguiendo este planteamiento hemos visto, por ejemplo, lo que se pudo averiguar acerca de los lazos fuertes y los lazos débiles, y tantas otras muestras de que sus posibilidades son de suma utilidad para establecer y descubrir relaciones entre nodos (personas, empresas, entidades, etc.) a través de vértices y en que todo ello nos permite ir más allá de lo que seríamos capaces sin este método de trabajo. Para nuestra investigación, destacamos el apunte ya hecho acerca de la capacidad de la Teoría de Grafos para indicarnos personas con mayor influencia y facilidad de acceso a la información. A la hora de realizar la investigación en Twitter con el fin de demostrar la posibilidad de hallar *insights* a través de las conversaciones entre sus usuarios, este tipo de posibilidades que nos ofrece la Teoría de Grafos serán las que guíen gran parte de nuestra labor. En este punto ampliaremos y desarrollaremos las posibles aplicaciones que nos convienen y queremos remarcar de cara a profundizar en las herramientas seleccionadas para la investigación.

Como herramienta privilegiada para el estudio de las relaciones que tienen lugar dentro de la sociedad, se establecen las “redes” (concepto cuyo estudio sistemático asentaba J. C. Mitchell a finales de los años sesenta). Básicamente, una “red” es un grafo que contiene, en cada uno de sus nodos (y/o aristas), información adicional (como ampliaremos en los próximos

párrafos. A nivel académico, es Harary el que expresa la adecuación de La Teoría de Grafos al análisis de las relaciones sociales: *“La teoría de grafos es un instrumento matemático adecuado para el análisis de las redes de relaciones sociales porque éstas son, como hemos visto, relaciones binarias definidas en conjuntos finitos de objetos. Y los grafos, en su definición formal son, precisamente, relaciones definidas sobre conjuntos finitos no vacíos”*. (Harary, 1969:7)^[196]

Asimismo, de la red social podemos extendernos al estudio de las redes de comunicación: cualquier red de comunicación puede representarse mediante un grafo. Cada posición en la red corresponde a un punto en el grafo y cada vínculo de comunicación corresponde a una arista o línea que conecta a un par de puntos. (Freeman, 2004)^[190]

Del mismo modo que un concepto puede expresarse en un grafo, la relación entre dos conceptos también puede serlo. Un ejemplo podría ser la relación “casado con”. En este grafo, los dos conceptos que se presentan son los sujetos casados entre sí. Si “Juan” está casado con “María”, tanto el concepto “Juan” como “María” tienen que ocurrir en este grafo que representa la sentencia. Esta unión de conceptos en un grafo permite manejar otro nivel de relaciones sociales.

Adentrándonos en el plano del Análisis de Redes Sociales ARS, recordaremos que tiene múltiples aplicaciones en campos muy diversos de las ciencias y la ingeniería; la computación, las matemáticas aplicadas (la teoría de los grupos finitos, la teoría de probabilidades, el análisis numérico y la teoría combinatoria), la teoría de la optimización, la planificación estratégica, entre otros. Puede servir también para la modelización de sistemas, la simulación, la estructuración de datos y el análisis y diseño de algoritmos. (Foulds, 1992)^[163]

3.3.1. Aplicaciones por sectores de actividad

Por su versatilidad, mostrada en repetidas ocasiones a lo largo de este punto de la tesis, la Teoría de Grafos es ampliamente utilizada para estudiar y desarrollar diferentes aplicaciones en múltiples áreas. Su implantación abarca desde el estudio de las moléculas a los enlaces químicos y hasta los átomos. En cuanto a campos del saber, su práctica abarca desde las ciencias sociales a las naturales. En el campo de la sociología, permite medir el prestigio de famosos (y personas influyentes en cualquier ámbito mediático) o los mecanismos de difusión que poseen las diversas redes sociales y sus nodos. En el estudio de la biología es útil para analizar la conservación de las especies, sus rutas migratorias o la propagación de las enfermedades. En cuanto a la aplicación en terrenos propiamente empresariales, puede ser usado en logística y transporte para la optimización de las rutas, identificando las rutas más cortas entre dos vértices de un grafo, por ejemplo. (Shirinivas, 2010:4611)^[233]

3.3.2. Tipos de redes

Dentro de la tipología de redes sobre las que puede desplegarse el análisis mediante la Teoría de Grafos, las redes más estudiadas son las redes tecnológicas, las redes sociales, las redes de información y las redes biológicas. Hemos considerado oportuno clasificar las redes por categorías sabiendo que alguna de ellas puede pertenecer a diversos grupos. (Newman, 2010:34)^[234]

3.3.2.1. Redes tecnológicas

Internet⁴⁸ es una de las más importantes de esta tipología. Internet se concibe como una red mundial de conexiones de datos entre ordenadores y dispositivos relacionados. Su representación más sencilla en la Teoría de Grafos sitúa en los vértices (nodos) de la red los ordenadores y otros dispositivos y en las aristas las conexiones entre ellos (Newman, 2010:35)^[234].

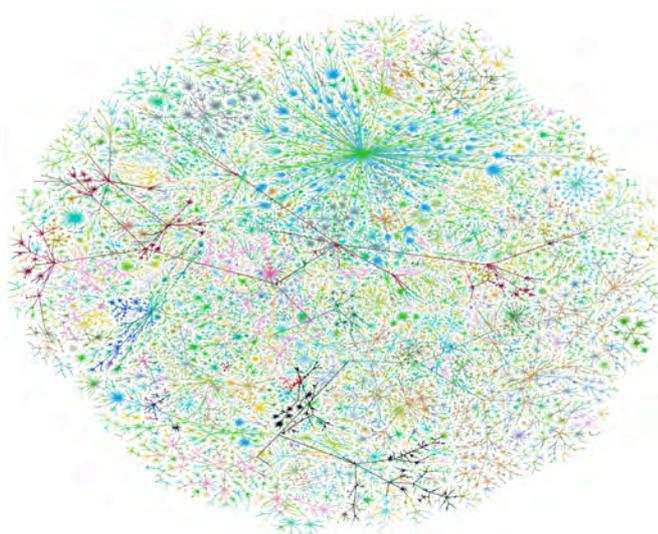


Figura 26: Estructura de Internet al nivel de sistemas autónomos

Los vértices de esta representación de Internet son sistemas autónomos y las aristas muestran las rutas tomadas por los datos que viajan entre ellos^[235].

Más allá de la esfera digital, las redes tecnológicas que tejen los sistemas de transporte son materia de intenso estudio con la aplicación de la Teoría de Grafos. Se conoce a Pitts como el pionero en su campo, cuando llevó a cabo un estudio del tráfico fluvial en los ríos rusos (Pitts, 1965)^[236]. Más recientemente, varios autores han presentado análisis de carreteras, ferrocarriles, vías de transporte, distribución logística y rutas aéreas. En la mayoría de los casos, los nodos representan ubicaciones geográficas y las aristas las rutas entre ellos. (Amaral, 2000)^[237] (Sen, 2003)^[238] (Gastner, 2004)^[239] (Easley, 2010:27)^[219]

⁴⁸ Internet es una red de datos en la que los mensajes enviados a través de ella se dividen en paquetes (pequeños “trozos” de datos) y se envían por separado para volverse a unir en un mensaje completo al alcanzar el otro extremo (Newman, 2010:35)^[234].



Figura 27: Ejemplo de red de rutas aéreas en la que los nodos representan los destinos y las aristas representan las conexiones directas. (Maps, 2015)^[240]



Figura 28: Ejemplo de mapa del metro de Washington. (Metrail, 2014)^[241]



Figura 29: GPS, ejemplo paradigmático de la aplicación de la Teoría de Grafos, su utilidad esencial es encontrar el camino más corto para alcanzar un destino señalado. (Garmin, 2014)^[242].

3.3.2.2. Redes sociales

En el ámbito de la ciencias sociales un dígrafo o grafo dirigido se usa, habitualmente para ilustrar las relaciones entre un conjunto dado de objetos. Cada objeto queda representado por un vértice (nodo) y la relación entre varios pares de objetos se representa, o bien por un arco si la relación es ordenada (por ejemplo: más alto que) o bien por una arista si la relación no está ordenada (por ejemplo: casado con) (Foulds, 1992:197)^[163]. Los investigadores de las redes sociales han desarrollado su propia terminología, substituyendo, por ejemplo, “nodos” por *actores* y “aristas” por *enlaces* (en inglés: *links*) (Newman, 2010:56)^[234]. Si nos adentramos en campos concretos de las ciencias sociales, la Teoría de Grafos demuestra ser particularmente útil en lingüística (Milroy, 1985)^[243] en el desarrollo del Procesamiento del Lenguaje Natural (*Natural Language Processing NLP*), (Turing, 1950)^[244] y en otros ámbitos de la lingüística como el *document clustering*⁴⁹ (Zhong, 2005)^[245] (Zha, 2002)^[246], el *Word sense disambiguation*⁵⁰ (Agirre, 2007)^[247] y el *Word clustering*⁵¹. (Matsuo, 2006)^[248]

⁴⁹ *Document clustering*: técnica para la organización no supervisada de documentos, filtrado y recuperación automática de contenidos. Se aplica en los motores de los buscadores que agrupan automáticamente los documentos más afines al objeto de búsqueda.

⁵⁰ *Word Sense Disambiguation WSD*: estudia el significado de una palabra dentro de una frase, cuando una misma palabra puede llegar a tener múltiples significados (Stevenson, 2005:249)^[253].

⁵¹ *Word Clustering*: técnica que consiste en dividir conjuntos de palabras en subconjuntos de palabras semánticamente similares. Aplicada en la Procesamiento del Lenguaje Natural (PLN). (Saha, 2008)^[254]

Hemos apuntado con anterioridad que, en el campo de la sociología, el concepto más utilizado es el de “Capital Social” (Portes, 2000:44)^[249]. Una de las utilidades de la Teoría de Grafos en esta disciplina es el cálculo de la relevancia, el prestigio o la influencia de las grandes personalidades (Katz, 1953)^[250] (Rusinowska, 2011)^[251] y, otro no menos relevante, el estudio de la difusión de la información, desde el cómo se propaga a quién posee la mayor influencia en su transmisión. (Bakshy, 2012)^[252]

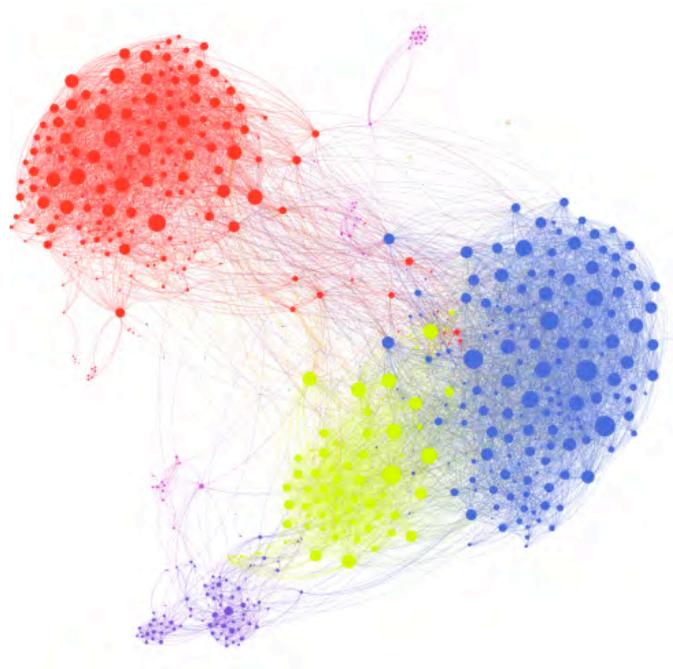


Figura 30: Ejemplo de una red de amistades en Facebook (Grafo no dirigido) (Griffgraphs, 2014)^[255]

3.3.2.2.1. Análisis de Redes Sociales (ARS)

Nos centraremos aquí con mayor detenimiento en el Análisis de Redes Sociales ARS: dicho método analítico se ha configurado en la actualidad como base para un gran número de estudios sobre la materia y, como explicaremos a continuación, consideramos que es la herramienta idónea para nuestro cometido: la investigación en Twitter con el objetivo de hallar *insights* publicitarios.

Del inglés: *Social Network Analysis* (SNA), el ARS surge como convergencia de varias corrientes de pensamiento (análisis antropológico), teorías de las ciencias sociales (la tradición “Gestalt” en psicología) y del uso de las matemáticas. (Scott, 1991:11)^[166]

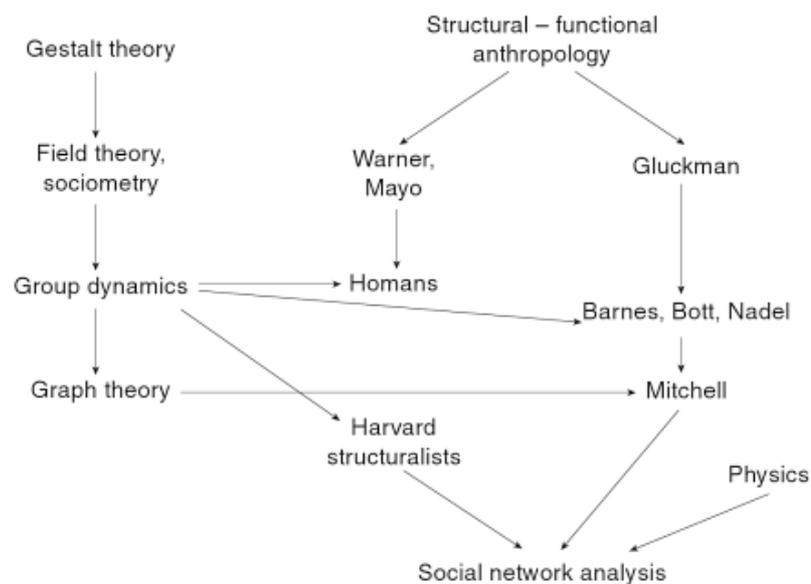


Figura 31: Linaje del Análisis de Redes Sociales ARS. (Scott, 1991:12)^[166]

El ARS, del mismo modo que se aplica la Teoría de Grafos en las redes sociales en general, estudia las relaciones sociales como patrones con “nodos” y “aristas” en un espacio matemático y cuyas propiedades formales son analizables con precisión. Aplicando la terminología concreta, los nodos son los “actores individuales” dentro de las redes y las aristas corresponderían a las relaciones entre los individuos (Easley, 2010)^[219]. En cuanto a las investigaciones en las suele ser más utilizado, el ARS se aplica al análisis de las

relaciones entre personas, grupos de personas, amistades, la transmisión de enfermedades, las relaciones sexuales y, más allá de todo ello, a cualquier relación humana que pueda ser segmentada en sujetos y vínculos (Pinheiro, 2011)^[256]. En la misma línea, el ARS ha sido utilizado en epidemiología para ayudar a comprender los patrones del contacto humano o para inhibir la propagación de una enfermedad entre la población. En este sentido, la evolución de las redes sociales puede ser modelada a partir del uso de otros modelos basados en agentes y que proporcionan información sobre la interacción entre las normas de comunicación, la propagación de rumores y las estructuras sociales. (Stattner, 2011)^[257]

El ARS también puede ser una herramienta eficaz para la vigilancia masiva; por ejemplo, se ha utilizado en el programa norteamericano denominado “*Total Information Awareness program*”. Con ello, el gobierno norteamericano estudió las relaciones y tendencias políticas junto con otros rasgos de la población norteamericana para determinar si algunos de sus ciudadanos podían suponer una amenaza. (Murray, 2010)^[258]

Para adentrarnos en la elaboración de un estudio aplicando el ARS, nos situaremos ahora en la investigación de James Fowler y Nicholas Christakis. En su conclusión, demostraron que la felicidad guarda una correlación en las redes sociales. Los investigadores resaltaron que, cuando una persona es feliz, sus amigos cercanos tienen hasta un 25% más de probabilidad de ser felices. También apuntaron que las personas que están en el centro de una red social tienen más posibilidades de ser felices en un futuro que los que están en la periferia. Los segmentos realizados de “personas felices” y “personas infelices” los situaron en un alcance de tres grados de separación, es decir, partiendo de que el grado de felicidad de una persona puede relacionarse con el de los amigos de los amigos. (Fowler, 2008:4)^[259]

El ARS, visto como un enfoque de investigación social, presenta las siguientes premisas fundamentales: (Freeman, 2012:3)^[178]

- El análisis de las redes sociales parte de la intuición estructural de la existencia de lazos que ligan a determinados actores sociales.
- Está basado en el estudio de información empírica sistemática.

- Hace un amplio uso de imágenes gráficas
- utiliza modelos matemáticos y/o computacionales.

Efectivamente, cumpliendo las premisas señaladas, el estudio de las redes sociales con el ARS permite comprender comportamientos, actitudes y otras manifestaciones que pueden reportarnos información muy valiosa para la toma de decisiones. En su despliegue, el ARS ha desarrollado una metodología para analizar las relaciones sociales, especializándose y profesionalizándose con planteamientos propios de toda índole: teóricos, metodológicos o con la programación de software analítico.

Para nuestra investigación resultará crucial la minería de textos (MT) o *Text Mining* (Definición en Terminología), la herramienta que permite obtener los textos producidos en las aplicaciones de la web 2.0 para su posterior estudio. Podremos establecer tipos de relaciones entre personas, señalar las más populares e influyentes, encontrar comunidades y establecer jerarquías. Este trabajo nos permite adentrarnos perfectamente en el terreno de las opiniones de los usuarios en Twitter acerca de sus intereses y su actividad diaria. Hallaremos ya actitudes respecto a determinados temas, evidencias de cierto tipo de relaciones sociales y todo ello, trabajado con detenimiento, nos dará una información valiosa para incorporar al trabajo realizado para generar *insights* publicitarios. (Bohn, 2009)^[260]

Por su notable complementariedad a la hora de extraer y relacionar datos diversos, se han producido interacciones entre el ARS y el MT. Es conveniente distinguir entre combinaciones para enriquecer la MT a través de la ARS de aquellas que lo hacen a la inversa. Desde el MT, una de las aplicaciones más célebres es la que incorporan los motores de búsqueda (Google o Yahoo!, por ejemplo). Los buscadores, entre otros criterios, clasifican los resultados de las búsquedas usando la herramienta MT a partir de medidas de centralidad basadas en grafos de hipervínculos. (Brin, 1998)^[261]

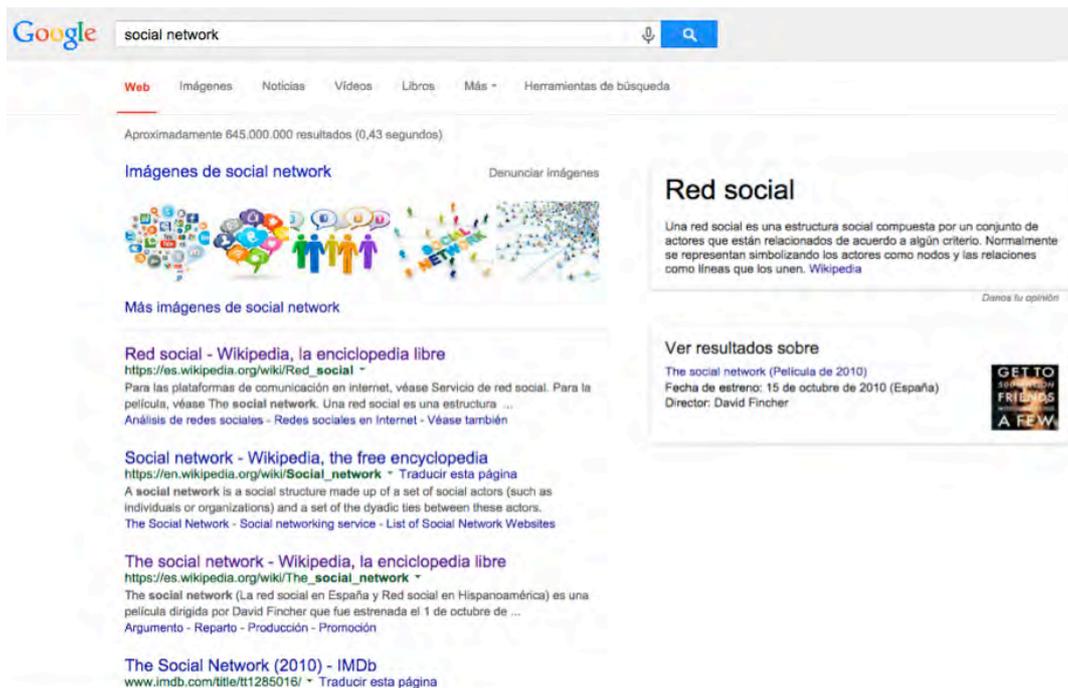


Figura 32: resultado de una búsqueda mediante una palabra clave en Google. (Google, 2014)^[262]

Otra aplicación destacada en la combinación del MT y el ARS es el resumen de textos: realiza el cálculo de las medidas de centralidad en redes de palabras (los nodos representan palabras conectadas si aparecen dentro del mismo documento de texto). Un ejemplo popular en la clasificación de contenidos en Internet es el que hace referencia a las nubes de palabras clave (*tag cloud*). (Bohn, 2009:13)^[260]

3.3.2.3. Redes de información

De todas las redes de información, la más conocida en la actualidad es la *World Wide Web*. Sin embargo, existen otras redes de información interesantes, como las redes de citación de varios tipos u otras como las redes que operan por correo electrónico. En el campo de corporaciones, hallamos redes profesionales de los servicios de redes sociales como Facebook y LinkedIn, así como las redes de *weblogs* o revistas digitales. (Newman, 2010)^[234]

La *World Wide Web* (Web):

Si entendemos la Web desde la Teoría de Grafos, la veremos como una red compuesta de nodos (sitios web) compuestos por textos, imágenes y otras informaciones y de aristas (hipervínculos) que nos permiten navegar de un sitio web a otro. Ya que los hipervínculos van en una sola dirección, la Web es una red dirigida.



Figura 33: representación de una red de páginas de un sitio web corporativo. Los nodos representan páginas del sitio web y las aristas dirigidas representan los hipervínculos existentes entre las páginas. (Newman, 2010:88)^[234]

Redes de citación:

Las redes de citación constituyen una antigua red de información vinculada a la red de citaciones de documentos académicos. La mayoría de los documentos académicos hacen referencia a otros documentos y ello queda representado por una bibliografía al final del documento. Efectivamente, se puede construir una red en la que los nodos sean los documentos trabajados y en el que las aristas estén dirigidas del documento A al documento B (si A cita B en su bibliografía). (Newman, 2010)^[234]

3.3.2.4. Redes biológicas

Las redes son ampliamente utilizadas en el campo de la biología como una representación de los patrones de interacción entre diferentes elementos biológicos. Los biólogos moleculares utilizan las redes para representar los patrones de reacciones químicas. Por su lado, los neurocientíficos las utilizan para representar los patrones de conexiones entre las células cerebrales. Destacamos también la utilidad en el terreno de la ecología: permiten las interacciones entre las especies dentro de los diversos ecosistemas. (Zhu, 2007)^[263]

Redes Bioquímicas:

Dentro de las redes biológicas, las bioquímicas son las que han obtenido mayor atención en los últimos años. Las redes bioquímicas permiten representar los patrones de interacción molecular y los mecanismos de control en la célula biológica. Las principales redes estudiadas en este área son las redes metabólicas, las redes de interacción proteína-proteína y las redes de regulación genética. (Newman, 2010)^[234]

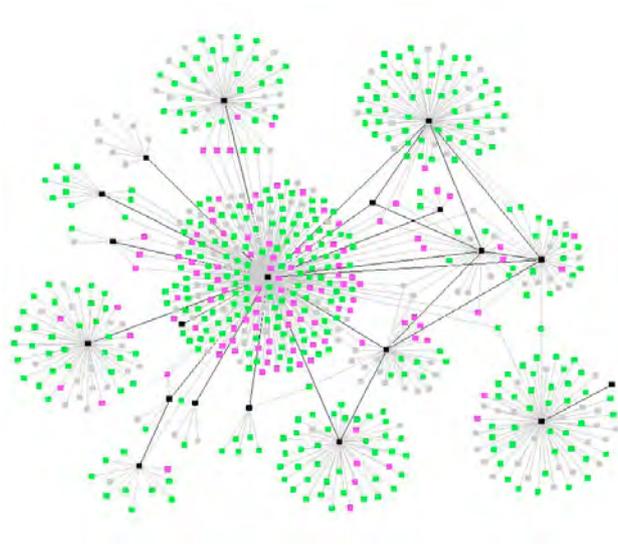


Figura 34: Red que representa la propagación de un brote de tuberculosis. (Andre, 2007:474)^[264]

Redes Neuronales:

El trabajo inicial de Watts y Strogatz (ver antecedentes) fue el primer caso de una aplicación de la Teoría de Grafos a una pregunta científica con el fin de estudiar la conectividad anatómica del sistema nervioso. (Watts, 1998)^[208]

Una de las aplicaciones del estudio de las redes en el terreno de la biología surge del estudio del cerebro y el sistema nervioso central. El cerebro procesa la información teniendo a las neuronas como primer elemento transmisor. Considerado el objeto más complejo del Universo, los intentos por entender el funcionamiento del cerebro y sus complicados patrones de conexión (tanto a en funcionamiento normal como en sus interacciones) es uno de los mayores retos de la ciencia moderna (Sporns, 2006:19219)^[265]. El análisis de grafos se ha utilizado en el estudio de modelos de redes neuronales y los resultados sugieren que el cerebro humano puede ser modelado como una red compleja y que puede tener una estructura similar a la que planteó *The Small World Problem*, tanto desde el punto de vista anatómico como desde el punto de la conectividad funcional. En la misma línea de investigación, se han hallado evidencias crecientes de que varios tipos de enfermedades cerebrales (Alzheimer, esquizofrenia, tumores cerebrales y la epilepsia) pueden estar asociadas con desviaciones de

la topología funcional de la red que puede establecerse partir de un patrón óptimo de “*Small World*”. (Stam, 2007:3)^[266]

Redes Ecológicas:

Las especies en un ecosistema pueden interactuar de muy diversas maneras. Aunque las diferentes formas de interacción se podrían representar en una sola red, tradicionalmente, los ecologistas separan los diversos tipos de interacción, dentro de las diversas relaciones, se han focalizado en especial en la relación depredador-presa. Pero existen muchas otras redes, como la que forman huéspedes y parásitos (Newman, 2003)^[267]. En concreto y como hemos apuntado con anterioridad, la Teoría de Grafos se aplica en el análisis de la conservación de las especies. Los vértices pueden representar regiones o hábitats en los que existen ciertas especies y las aristas pueden representar las rutas de migración o los movimientos entre regiones que realizan los animales. Con esta información, organizada como un grafo, se determinan patrones de reproducción o de propagación de enfermedades o parásitos. Ello tiene en cuenta que cualquier cambio en los desplazamientos de las especies puede afectar a su evolución. (Mason, 2007)^[268] (Bunn, 2000)^[269]

Como conclusión de este apartado constatamos que la aplicación de la Teoría de Grafos al estudio de las redes sociales goza de una situación privilegiada y que se nos presenta como idónea para nuestra investigación. Además de la multitud de campos y otras ciencias y disciplinas en las que puede aplicarse, las posibilidades que hemos presentado en cuanto a la capacidad de la TdG para detectar de forma medible el prestigio de ciertos usuarios, la identificación de palabras clave y enunciados más importantes (más allá de la cantidad, la clave es la medida que hemos presentado como “centralidad”) constituyen la raíz y la fuente de obtención de información que nos permitirá realizar la posterior investigación para hallar los *insights*. Como empresa del campo de la comunicación que ofrece servicios a través de redes sociales, con la fuerte presencia de usuarios *influencers* y de marcas que también utilizan el servicio, Twitter es un lugar idóneo para aplicar la TdG en el contenido de los tuits que emiten sus usuarios. Ello será posible gracias a técnicas como la definida anteriormente como Procesamiento del Lenguaje Natural. Presentaremos, en los capítulos siguientes de esta tesis, cómo hemos llevado a cabo de forma concreta y precisa la aplicación de la TdG en diversas técnicas del campo de las redes sociales.

4. RESUMEN AUTOMÁTICO DE TEXTOS

4. Resumen Automático de Textos (RAT)

A finales del siglo pasado ya se afirmaba que gracias al desarrollo tecnológico, en los últimos 50 años del siglo XX, se había puesto a disposición de las personas tanta información como nunca antes en la historia de la humanidad (Feather, 2008)^[270] e Internet había jugado un papel principal en la sobreoferta informativa desde su entrada en sociedad en los inicios de los años 90 del siglo pasado (Swash, 1998)^[271]. Los medios saturan de manera creciente a las audiencias con su oferta de información no siempre buscada por los individuos (Edmunds; Morris, 2000)^[272] de tal forma que la supervivencia personal y profesional en las sociedades modernas depende de la habilidad de discriminar entre la ingente oferta de información mientras ésta crece de forma exponencial. (Lewis, 1999)^[273]. (Del Fresno, 2012:105)^[66]

Antes de la existencia de internet y la expansión de su uso a escala doméstica hasta formar parte del día a día de miles de millones de personas en todo el mundo mediante los ordenadores (y, más recientemente, todo tipo de dispositivos móviles), la constancia “tangible” que quedaba de las conversaciones y de todo tipo de comunicaciones era mucho más limitada. El primer quiebro excepcional en el transcurso de la comunicación de masas vino con la invención de la imprenta a mediados del S.XIV. La imprenta de Gutenberg fue la primera herramienta que permitía la difusión exacta de un texto realizada de forma mecánica, más allá de la copia manual que se usaba hasta entonces o de la comunicación oral. Si bien el primer periódico del mundo, La Gazette de France, surgió en el S. XVI, la implantación de la prensa escrita como medio de comunicación de masas tardaría todavía varios siglos. Hasta la invención del teléfono y de los equipos de grabación, a mediados del siglo XXI, no existiría otra forma de conservar y reproducir con exactitud comunicaciones entre personas. En los albores del siglo XX la aparición de la radio sería el nuevo salto hacia un nuevo modo de comunicación de masas. Pronto llegaría la televisión y, de ese modo, se completaría lo que anteriormente hemos definido como medios de comunicación de masas tradicionales.

La cita que ha dado inicio a este capítulo ponía de manifiesto que, a mediados del siglo XX, ya se había puesto a disposición de las sociedades tanta información “como nunca antes en la historia de la humanidad”. Con la llegada de internet, el crecimiento supera de manera excepcional esa afirmación, dejándola prácticamente obsoleta. Según datos obtenidos en este

último año⁵², cada 60 segundos se crean 571 webs, se suben 20.000 fotos a Tumblr, 72 horas de vídeo a YouTube, 350 gigas de datos a Facebook y 278.000 tuits. Google recibe 2.000.000 de búsquedas, 20 millones de imágenes se visionan en Flickr y se envían 204 millones de correos electrónicos. A esta descomunal cantidad de tráfico se la ha denominado “Big Data”. (Snijders, 2012)^[274]

Las cifras acerca de la circulación de datos en la red siguen creciendo a diario y su repercusión es también inédita en comparación a la que se registraba antes de su existencia. Sin embargo, más allá de todo eso, uno de los mayores cambios en el mundo de la comunicación es que todo lo que ocurre a través de internet queda, de un modo u otro, registrado. Una parte considerable de la información (desde las webs comerciales a los medios de comunicación de masas, los blogs o las conversaciones en Twitter) es accesible públicamente.

El *boom* de internet ha ido de la mano del avance de las nuevas tecnologías y, por ende, del perfeccionamiento y la aparición de las herramientas que permite el acceso y estudio para diversos fines de la información que circula a través de internet, sea esta entre personas, empresas o entre el público y las marcas. Destacamos, de nuevo, el resumen automático de textos o el reconocimiento y transcripción de mensajes de voz. Las técnicas que se basan en el Procesamiento del Lenguaje Natural (PLN).

Al mismo tiempo, la tecnología ha ido avanzando, para permitir el análisis de dichas conversaciones a través del resumen automático de los textos, así como la capacidad, cada vez mayor, del reconocimiento de la voz y su transcripción, para ser procesado aplicando las técnicas del Procesamiento del Lenguaje Natural (PLN) (Ver definición en Terminología). (Carenini, 2011)^[275]

Una de las consecuencias directas de esta nueva situación es lo que Alfons Cornellà describe en su libro *Infoxicación* y que se resume en la imposibilidad e incapacidad de un ser humano

⁵² Recopilación de datos realizada por la empresa Qmee procedentes de www.pcmag.com, www-go-gulf-com, www.businessinsider.com, www.dailymail.co.uk, www.4mat.com y www.scoop.intel.com

para consumir, de forma ordenada y completa, el flujo de información generado acerca de un tema de su interés. (Cornella, 2000)^[276]

La información se disemina entre conversaciones, webs, blogs, foros de opinión, imágenes, vídeos y es imposible lograr reunirla en su totalidad.

Esta problemática se extiende también en el ámbito empresarial, en el que las grandes empresas destinan numerosos recursos a intentar rastrear la red en busca del contenido que se publica y comparte acerca de ella. Los líderes políticos y personajes públicos de gran relevancia disponen también de servicios que ordenan y clasifican en la medida de lo posible todo aquello que les concierne directamente. Esta posibilidad se puede alcanzar gracias a las tareas del Procesamiento Natural del Lenguaje (PLN) (Ver definición en Terminología), de entre las cuales hallamos el resumen automático de textos. A partir de un texto de origen, el contenido más importante se puede presentar de forma reducida y adecuada a las necesidades de cada usuario. Al igual que las marcas usan el Procesamiento Natural del lenguaje para cuidar su imagen y los políticos para medir la reacción que generan sus decisiones, los publicistas pueden usarlo para sus propios fines, uno de los cuales seguimos desplegando a continuación. (Spärck Jones, 1998)^[277] (Manning, 2000)^[191] (Hovy, 2005)^[278]

Como se verá en la tercera parte de esta tesis doctoral, emplearemos la herramienta de resumen automático de textos “SETAL”. La herramienta destaca por relacionar, resumir e identificar textos que expresen sentimientos comunicados a través de las plataformas de servicios de redes sociales digitales (como Facebook o Twitter). Para profundizar en el concepto y el funcionamiento del Resumen Automático de Textos, seguiremos con el mismo planteamiento que el realizado en el capítulo anterior con la Teoría de grafos y señalaremos sus antecedentes.

4.1. Antecedentes del Resumen Automático de Textos

Se sitúa el nacimiento del interés por el resumen automático de textos en el sistema bibliotecario norteamericano. Debido al creciente catálogo de las obras almacenadas, las bibliotecas de Estados Unidos acusaron la necesidad de indexar de forma digital sus contenidos y, ante las limitaciones de almacenamiento disponible en la época de 1950, tuvieron que idear un sistema capaz de resumir los contenidos de la forma más precisa posible. Inicialmente, la mayor parte de la labor de resumen se centró en documentos de carácter técnico. (Das, 2007)^[279]

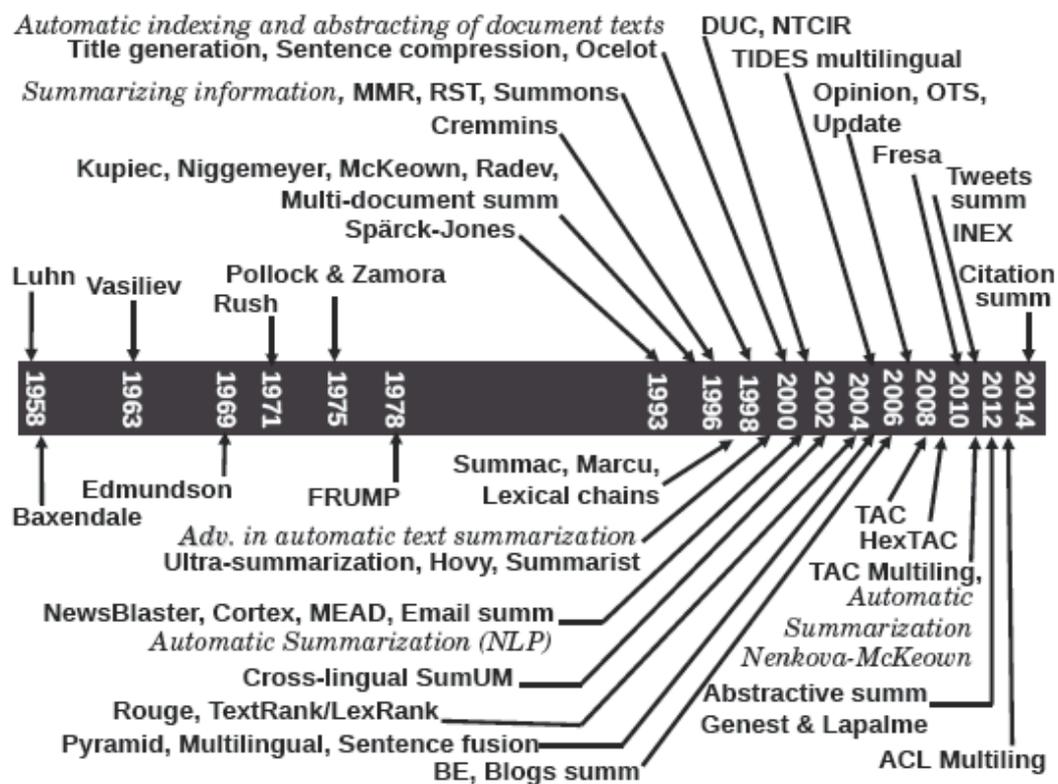


Figura 35: Cronograma de los autores y temas más relevantes de la evolución del resumen automático de textos. (Torres-Moreno, 2014:17)^[280]

Sin embargo, las primeras muestras de elaboración de resúmenes automáticos mediante computadoras y sin intervención humana, tienen lugar a finales de los años cincuenta. En este

sentido, Luhn describió las labores de investigación desarrolladas por la empresa IBM en el comienzo de este campo de investigación. (Luhn, 1958)^[281]

Las primeras investigaciones se centraron en la obtención de la frecuencia de uso de ciertas palabras, considerando que la frecuencia de una palabra dentro de un texto proporcionaba una medida útil de su contenido (Luhn, 1958)^[281]. Un trabajo relacionado con el de Luhn y desarrollado también desde IBM fue el que consideraba la posición de la frase u oración principal dentro del texto. En esta línea, Baxendale analizó 200 párrafos y descubrió que en el 85% de ellos la oración principal aparecía al inicio del texto y la frase final en un 7% (Das, 2007)^[279] (Lin, 1997)^[282]. Dos autores de suma relevancia en aportar diversos avances al RAT han sido Edmunson, en cuanto al trabajo con la detección de las “palabras clave” del texto, y Salton, cuyo trabajo con la denominada *Bag of Words* ha permitido, entre otros avances, el desarrollo de las “nubes de *tags*”. (Edmundson, 1969)^[283] (Salton, 1975)^[284]

Como se observa en el cronograma, en los años ochenta decrecieron las investigaciones en la materia y no sería hasta la década siguiente que se daría un resurgimiento protagonizado por grandes progresos. Los avances vinieron encabezados por las invenciones en el campo de la Inteligencia Artificial, la utilización de sistemas híbridos y la aparición de las técnicas del aprendizaje automático (*Machine Learning*) en cuanto al PLN. Asimismo, en la misma década empezaron los programas de investigación basados en técnicas estadísticas aplicadas al resumen automático de textos su evolución ganó complejidad y se aplicó en dos grandes áreas: la estadística y la lingüística. (Salton, 1988)^[285] (Kupiec, 1995)^[286]

En las investigaciones llevadas a cabo en el campo estadístico se aplicaron los modelos bayesianos (Kupiec, 1995)^[286], las técnicas de *clustering* (Radev, 2000)^[287], los grafos (Leskovec, 2005)^[288] y el *Machine Learning*, entre otros (Berger, 2000)^[289] (Barzilay, 2008)^[290]. En cuanto al grupo de investigación lingüística, destacaron los trabajos que explotaron las posiciones textuales (Brandow, 1995)^[291] (Lin, 1997)^[282], la estructura del discurso, (Marcu, 1997)^[292] (Polanyi, 2004)^[293] y las cadenas léxicas (Barzilay, 1999)^[294] (Fuentes, 2008)^[295].

Como hemos señalado, en los últimos años, la existencia del “*Big Data*” no ha hecho más que acelerar la necesidad de mejorar las técnicas de resumen automático de textos. La invención de ordenadores con procesadores más rápidos y potentes ha contribuido a evitar el estancamiento y, de hecho, fruto de ello los avances han sido notables en cuanto al volumen de datos que se pueden llegar a filtrar con algoritmos más eficaces.

Ampliaremos en el siguiente punto la definición del RAT incorporando las diversas contribuciones de sus teóricos destacados.

4.2. Definición

Como hemos apuntado anteriormente, el RAT es un proceso de creación de un resumen de uno o más textos en el que un ordenador sintetiza un texto largo en una forma más corta y no redundante⁵³ ⁵⁴. Según Hovy, el resumen no debería ser más extenso que la mitad de los originales. Al resumen que obtiene el RAT le llamaremos *abstract*, que se caracteriza por indicar y predecir la estructura y el contenido del texto. (Mani, 1999)^[296] (Das, 2007)^[279] (Steinberger, 2009)^[297] (Torres, 2014)^[298] (Hovy, 2005:584)^[278]

Mani ofrece otra definición más sencilla pero igualmente ilustrativa: “A partir de la obtención de una fuente de información, extraer el contenido de ella y presentar el contenido más importante para el usuario de forma condensada y útil a las necesidades de aplicación del usuario”. (Mani, 2001, Introducción)^[299]

Desglosaremos en los tres puntos siguientes el esquema básico de funcionamiento de los sistemas RAT que siempre se repite: (Spärck Jones, 1998:2)^[277]

- Interpretación o análisis: se estudia el documento fuente y se construye una representación del mismo.
- Transformación: corresponde al resumen abstractivo y a las técnicas de PLN. Transforma la representación del documento en una nueva representación.
- Generación o síntesis: a partir de una representación del resumen, construye un nuevo resumen en lenguaje natural.

⁵³ “An abbreviated, accurate representation of the content of a document preferably prepared by its author(s) for publication with it. Such abstracts are also useful in access publications and machine-readable databases” (American National Standards Institute Inc., 1979)^[369]

⁵⁴ “A summary as a reductive transformation of source text to summary text through content reduction by selection and/or generalization on what is important in the source”. (Spärck Jones, 1998:2)^[277]

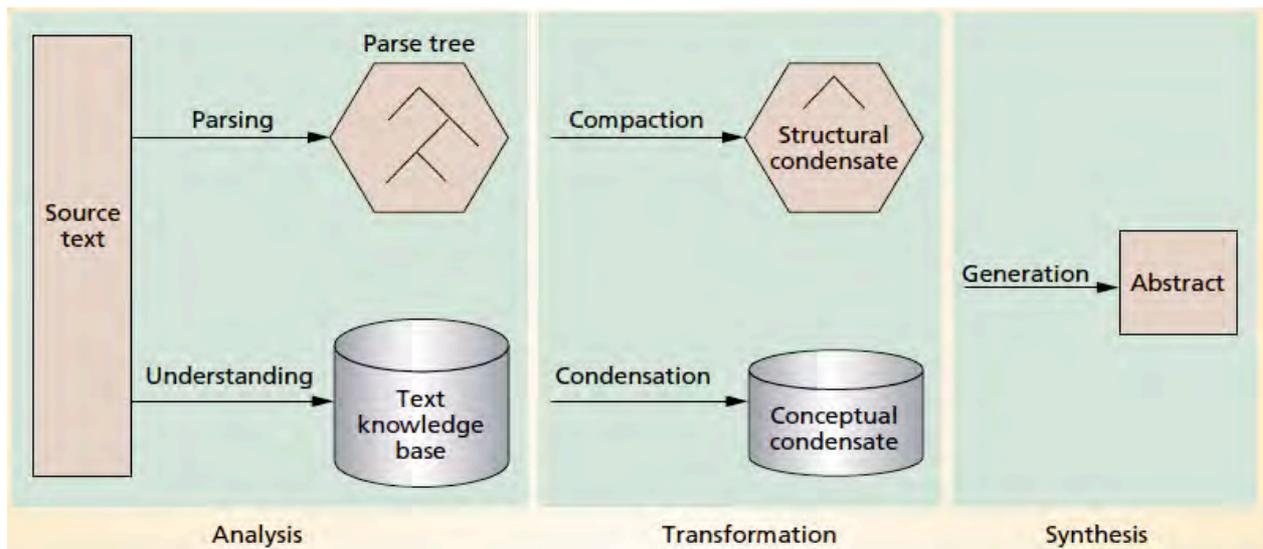


Figura 36: esquema que ilustra el proceso de abstracción desde dos enfoques básicos que a menudo se combinan. (Hahn, 2000:32)^[300]

El resultado de un *abstract* presenta las siguientes ventajas que constituyen la principal razón de su uso en nuestra investigación: (Borko, 1975)^[301]

- Promueven el conocimiento actual
- Ahorran tiempo de lectura
- Facilitan la selección posterior del contenido que consideraremos relevante
- Facilitan las búsquedas dentro de la lectura
- Mejoran la eficiencia en la indexación de contenidos
- Ayudan en la elaboración de las opiniones que utilizaremos posteriormente

Aunque los RAT pueden estar caracterizados por múltiples factores, son principalmente tres los que mayor influencia tienen: la entrada, el propósito y los factores de salida. Definiremos sintéticamente cada uno de ellos.

a) La entrada

Tradicionalmente y como su nombre indica, los RAT se basan en texto escrito, aunque también pueden realizarse a partir de información multimedia (imágenes, vídeos o audios. La proliferación en la red de los contenidos multimedia ha generado la necesidad de mejorar y renovar la forma de analizar los contenidos de información entrantes y ha establecido las siguientes relaciones junto a los textos asociados:

- Documento: único, múltiples
- Idioma: monolingüe o plurilingüe
- Genero: técnico, científico, noticias
- Longitud: corto (1 a 2 páginas), largo (> 50 páginas)
- Medio: texto, audio, video

b) El propósito

Como resulta patente, el propósito es “resumir” el contenido de los documentos y se puede subdividir en tres categorías según la situación (el contexto en el que será utilizado), el público (la clase de lector que hará el resumen) y el uso (como la recuperación de los textos originarios, las vistas previas o una sinopsis de un ciclo de conferencias, por ejemplo).

c) La salida

Existen, como mínimo, tres formatos de salida de la información: el material, el formato y el estilo. Con la salida “material” nos referimos a la extensión con la que el resumen pretende capturar todo el contenido importante del texto original o, en vez de ello, quizás algunos aspectos relevantes del mismo. Respecto al “formato” encontraremos gran variedad de formatos de salida (resúmenes de prensa etiquetados por categorías temáticas o resultados de tests biológicos, entre otros). El estilo puede ser informativo (centrado en una temática

contenida en el texto), indicativo (indicando el tema general sin dar más explicaciones), crítico (analizando el contenido del documento) o agregado (incorporando diferentes fuentes y de uso frecuente en el ámbito judicial). Esta lista la situamos a modo orientativo y general a sabiendas de que puede ser más exhaustiva. Sin embargo, para nuestro propósito, resulta suficiente. (Spärck Jones, 1998)^[277]

4.3. Taxonomía y métodos

Una de las ideas que hemos ido mostrando a lo largo de los anteriores apartados de esta tesis es la de hacer patente el enorme desarrollo del RAT a raíz de la revolución tecnológica reciente y de la masificación de internet. La posibilidad de obtener información valiosa de los contenidos que los usuarios comparten en la red, junto con la mejora de las herramientas disponibles, ha hecho posible que, por ejemplo, el RAT haya evolucionado de una técnica de recuento de palabras frecuentes a una técnica capaz de identificar, analizar y clasificar los sentimientos. Todo ello, efectivamente, gracias al *Machine Learning* y la Inteligencia Artificial.

En nuestra tesis aprovecharemos los mayores avances en la materia, tanto teóricos como técnicos, para proceder satisfactoriamente con nuestra investigación y, más allá de eso, con la voluntad de lograr extraer un método válido y concreto para la obtención de *insights* publicitarios.

4.3.1. Taxonomía de los métodos del RAT

Proponemos a continuación las descripciones de los diversos tipos de RAT más comúnmente utilizadas por los investigadores (Ježek, 2008)^[302]. Tal y como se ilustra en la figura 37, los RAT pueden obtenerse desde diversos planteamientos. De entre todos ellos, los más relevantes son los extractivos/abstractivos, los indicativos/informativos/críticos, los genéricos/*query-based*/guiados y, finalmente, los mono/multidocumento. Asimismo, pueden realizarse RAT con una aproximación idiomática mono/multilingüe. Definiremos a continuación cada uno de ellos.

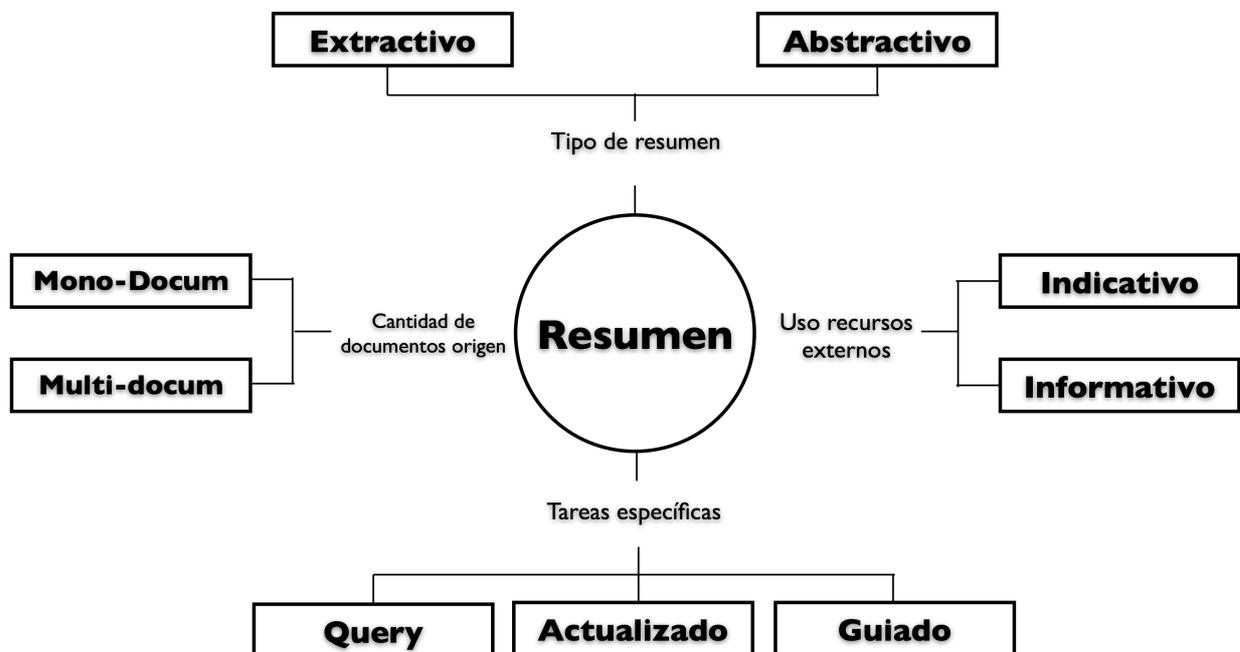


Figura 37: Principales aproximaciones para iniciar el proceso RAT. Adaptación propia a partir de (Sizov, 2010:8)^[303]

4.3.1.1. RAT según la forma

Dichos métodos se caracterizan por la forma en que se realiza el resumen. Por forma entendemos la manera en la que se obtendrá el resumen.

4.3.1.1.1. Resumen extractivo

Consiste en la obtención de secuencias de palabras realizada a partir de otras palabras, frases, oraciones o párrafos de texto que conforman un subgrupo dentro del texto original (Ježek, 2008)^[302]. Las frases extraídas se pueden comprimir, simplificar o agregar de diversas maneras. (Gupta, 2010)^[304]

4.3.1.1.2. Resumen abstractivo

En un resumen abstractivo la fuente de información no se encuentra de forma literal en el texto originario. Por ello, el método es más complejo y requiere técnicas de PLN más sofisticadas.

El contenido más relevante se selecciona y se expresa en la forma más parecida posible a lo que sería el lenguaje natural. La forma en la que un persona realizaría el resumen de un texto. El método abstractivo, además de un módulo de generación del lenguaje natural, requiere un sistema capaz de procesar y jerarquizar el conocimiento extraído. Se trata de crear un resumen a partir de la interpretación de la fuente de información original, generando un texto que exprese la misma información pero de una forma más concisa, resumida. (Das, 2007)^[279]

4.3.1.2. RAT según el propósito

4.3.1.2.1. Resumen indicativo

El objetivo del resumen RAT indicativo es realizar resumen para que los lectores comprendan la idea básica del texto original y puedan considerar si merece o no una lectura del total del documento. Su longitud habitual oscila entre el 5% y el 10%. del texto original. (Ježek, 2008:3)^[302]

4.3.1.2.2. Resumen informativo

La misión del RAT informativo es resumir la información más relevante del texto inicial pero, a diferencia del resumen indicativo, con la expresa voluntad de sustituirlo. Para lograr esta “sustitución” y mantener el sentido original sin recortar en exceso su contenido, los resúmenes informativos suelen tener mayor extensión y alcanzan entre un 20% y un 30% del texto original. (Ježek, 2008:3)^[302]

4.3.1.2.3. Resumen crítico

El resumen crítico, como variación del resumen informativo, incorpora la opinión e interpretación por parte de los encargados de realizarlo. (Hahn, 2000)^[300]

4.3.1.3. RAT según la audiencia

4.3.1.3.1. Genérico

El RAT genérico es el que va dirigido a un público objetivo o audiencia amplios y los temas principales se presentan con el mismo nivel de importancia. (Ježek, 2008)^[302]

4.3.1.3.2. Query-based

El RAT *query-based* se realiza a partir de las necesidades de un determinado usuario o colectivo y consiste en el análisis del contenido del texto original en busca de hallar datos que respondan a una pregunta (*query*) concreta. (Hovy, 1999)^[305] (Dang, 2008)^[306]

4.3.1.3.3. RAT “guiados” hacia un objetivo

Los resúmenes “guiados” no están totalmente automatizados y para llevarlos a cabo se requiere situar previamente una serie de variables. Por ejemplo, si una compañía de seguros quiere realizar un estudio de siniestros de una tipología de seguros y no de todos los seguros en general, deberá definir una serie de variables tales como la descripción del suceso, la fecha o la ubicación, entre otras (Nenkova, 2011)^[307]. Por su propio modo de elaboración, los RAT guiados llevan a cabo análisis semánticos más profundos que los que presentan estadísticas de frecuencias. Por su mayor complejidad en la elaboración y gracias a que la automatización de resúmenes recibe constantes mejoras, la tendencia es que los RAT guiados sean cada vez menos necesarios. (Dang, 2010)^[308]

4.3.1.4. RAT según la cantidad de documentos a resumir

Los resúmenes hechos a base múltiples documentos solucionan la necesidad de obtener información que se encuentra repartida en diferentes fuentes y de la que se quiere obtener una visión global. Las posibilidades de los RAT que combinan varios documentos son múltiples y habituales, desde la elaboración de una cronología de acontecimientos hasta las comparativas de artículos o los *clipping* de prensa. (McKeown, 1999)^[309]

4.3.1.5. RAT según el idioma

4.3.1.5.1. Plurilingüe

Las fuentes de información pueden ser diferentes al idioma del resumen final.

4.3.1.6. RAT según las fuentes

4.3.1.6.1. Multimedia

Realizados con contenidos de diverso formato, tales como vídeos, imágenes, grabaciones de audio, noticiarios televisivos o cualquier combinación entre ellos (Evans, 2006)^[310]

4.3.1.7. RAT según sus métodos

En los antecedentes al RAT hemos fijado los métodos más relevantes anteriores a su aparición, especialmente aquellos que buscaban extraer frases o palabras para comprender la temática o los aspectos más relevantes del contenido sin la necesidad de tener que abordar su lectura completa.

4.3.1.7.1. Métodos estadísticos

El método estadístico parte de la demostración de que el número de veces en las que aparece un término repetido en un texto guarda una relación con su importancia. A mayor repetición del término, mayor importancia tendrá en el desarrollo del contenido (Salton, 1988)^[285]. Como ejemplo paradigmático, la aparición de un *Trending Topic* (TT) en Twitter (selección de un *hashtag* representado por “#” y que acompaña a un término que elige el usuario y que refiere al tema de su tuit, como por ejemplo #NuevoiPhone #EleccionesFrancia o #ChampionsLeague), tiene que ver en parte al número de usuarios que lo están utilizando. Es decir, si una determinada cantidad de usuarios utiliza un *hashtag* en concreto en un momento

preciso y alcanza una masa crítica este puede aparecer en Twitter como TT. En el algoritmo de Twitter para calcular un TT se utilizan además otros factores aparte de la frecuencia de uso.

Como parte del método estadístico, mencionamos el clasificador bayesiano. Su utilización permite calcular la probabilidad de que una frase del documento original sea incluida o no en el resumen. La aplicación de la fórmula bayesiana puede señalar, además, el número de mayúsculas, la longitud de la frase, su estructura y la posición de la frase respecto al total del texto (Kupiec, 1995)^[286]

Queremos señalar también la importancia que tiene la posibilidad de medir el significado de un término en vez de su frecuencia de aparición dentro del texto. Con la herramienta WordNet puede llevarse a cabo este análisis usando el recuento del término elegido pero sumándolo a la aparición de sus sinónimos o de otros términos que hayamos decidido asociarle. (Hovy, 1999)^[311] (Ježek, 2008)^[302]

4.3.1.7.2. Métodos basados en la conectividad del texto (*Text Connectivity*)

La riqueza del lenguaje escrito puede ser para el RAT una complicación añadida. Por ejemplo, la aparición de frases con expresiones anafóricas⁵⁵ dificulta la detección de frases que podrían tratarse del tema principal aunque no lo parezcan. Este tipo de tesituras han sido investigadas desde diversas vías.

El método de las “Cadenas Léxicas” (*Lexical chains*) utiliza WordNet para buscar relaciones de cohesión entre términos (frecuencia de uso, sinónimos, antónimos, hiperónimos⁵⁶ y holónimos⁵⁷) y forja cadenas con su asociación (Barzilay, 1999)^[294]. De planteamiento

⁵⁵ Anáfora. Relación que establece una palabra (generalmente un pronombre) con otra u otras aparecidas previamente en el discurso y que permiten determinar cuál es su referente. (RAE, 2015)^[317]

⁵⁶ En Semántica lingüística, se denomina hipersónico a aquel término general que puede ser utilizado para referirse a la realidad nombrada por un término más específico. (Wikipedia, 2015)^[318]

⁵⁷ La holonimia es una noción semántica que se opone a meronimia, del mismo modo en que se oponen el todo y la parte. Así, por ejemplo, bicicleta es un holónimo mientras que sillín, pedal y aro son merónimos. (Wikipedia, 2015)^[319]

similar, el método de la co-referencia de contenidos identifica si dos expresiones del lenguaje natural se refieren al mismo tema (Boguraev, 1999)^[312]

Dentro de los métodos *Text Connectivity* encontramos la Teoría de la Estructura Retórica (*Rhetorical Structure Theory*), basado en una serie de relaciones teóricas que conectan entre sí unidades de texto. Las relaciones se representan en forma de árbol y el resultado se utiliza como resumen. Las oraciones se sitúan en diversas posiciones dentro del árbol y, en función del peso que se les otorga, reciben un valor: las oraciones satélites tendrán valor 1 y las más cercanas al núcleo tendrán valor 0. (Ono, 1994)^[313] (Khan, 2005)^[314]

4.3.1.7.3. Métodos de Grafos Iterativos

Los algoritmos que se utilizan en los métodos de grafos iterativos, como son el HITS (Kleinberg, 1999)^[315] o el *PageRank* de Google (Brin, 1999)^[316], sirven para estudiar las estructuras de los enlaces y con ello encuentran una manera de jerarquizar los sitios web. También se utiliza con éxito en otras áreas del análisis de redes de citación, como es el caso de las redes sociales. (Ježek, 2008)^[302]

4.3.1.7.4. Método de la Lógica Difusa (*Fuzzy Logic*)

EL RAT obtenido aplicando el método *Fuzzy Logic* tiene en cuenta todas las características del texto que hemos mencionado, como son la frecuencia de palabras, la similitud entre palabras clave, las semejanzas de contenido de los títulos, la posición de las oraciones y otros atributos gramaticales. El método, de forma parecida al Grafo Iterativo, realiza la extracción y puntúa cada frase con un valor entre el 0 y el 1. El valor obtenido determina el grado de importancia de la oración dentro del resumen final. (Kyoomarsi, 2008)^[320]

Hemos prodecido a presentar la relación de los métodos más utilizados para el RAT. El algoritmo desarrollado para el resumen de los textos obtenidos de las conversaciones en Twitter, realizadas por los usuarios analizados para esta tesis, se ha utilizado una combinación de métodos como el resumen extractivo a partir de un método *Query based* en

la que el interlocutor guía el algoritmo en su búsqueda a partir de un diccionario de palabras clave del sector de actividad que se analiza. A partir de los métodos estadísticos de clasificación binaria permiten establecer la importancia de las palabras dentro de los textos analizados.

4.3.2. Técnicas de evaluación de resúmenes

Las técnicas de evaluación de RTA constituyen una tarea sustancial que nos permitirán determinar la calidad del resultado obtenido. En el caso de la recuperación o categorización de textos, existen sistemas de evaluación consolidados (como el *Recall* o el *Precisión*). Por contra, la evaluación de resúmenes extractivos o abstractivos requiere técnicas de elaboración más compleja. Clasificaremos las técnicas según si son intrínsecas o extrínsecas. (Jones, 1996:19)^[321]

La calidad de los resúmenes generados a partir de la evaluación intrínseca, realizada por personas, tiene un nivel de subjetividad evidente, pero los criterios básicos se establecen a partir de la calidad de los resúmenes en cuanto al documento resultante del resumen, su capacidad para informar, la legibilidad, la presencia de anáforas sin resolver así como la falta de conservación de la integridad de los entornos estructurados como listas o tablas. (Minel, 1997)^[322]

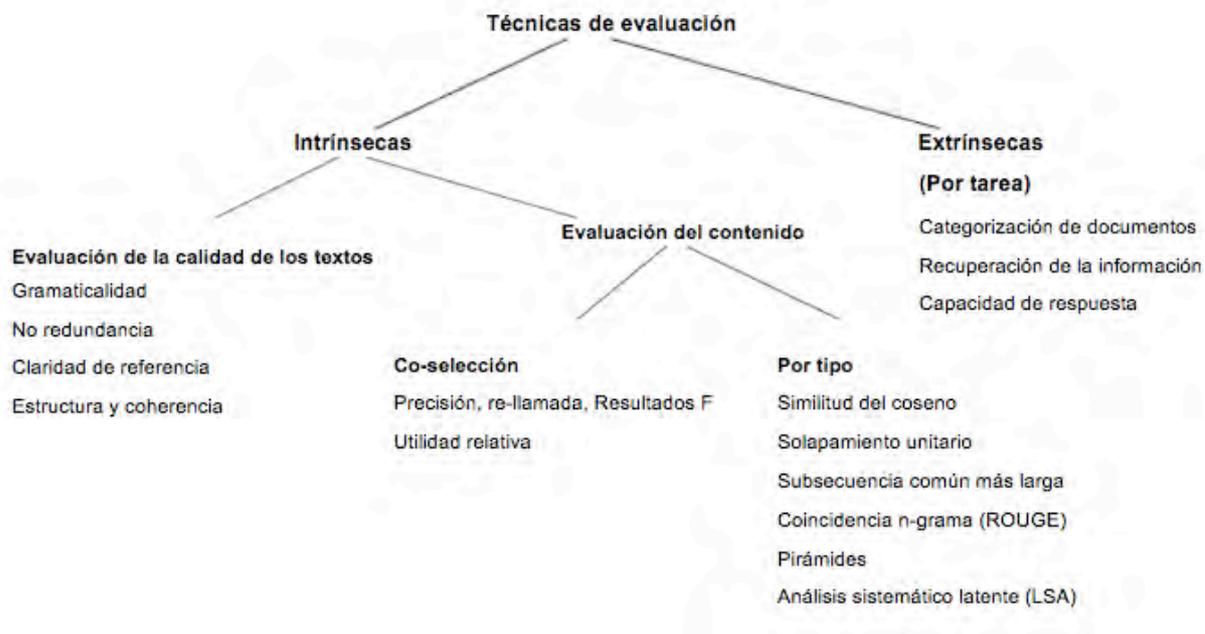


Figura 38: Principales técnicas de evaluación de RTA. (Steinberger, 2009:1006)^[297]

4.3.2.1. Evaluación Intrínseca

Las métricas intrínsecas evalúan el contenido de la información comparándolo o bien con otros RTA realizados por seres humanos o bien con la fuente original de la información. (Mollá, 2003)^[323]

4.3.2.1.1. Criterios a tener en cuenta para la evaluación de la calidad de los textos (*Text quality evaluation*)

Señalamos los aspectos a tener en cuenta desde el ámbito lingüístico: (Steinberger, 2009:13)^[297]

- Gramaticalidad: el texto no debe contener elementos no textuales (como marcadores, errores de puntuación o palabras incorrectas).
- No redundancia: el texto no debe contener información repetitiva.

- Claridad de referencia: los nombres y pronombres deben estar claramente referenciados. Por ejemplo, el pronombre *él* tiene que hacer referencia a alguien concreto dentro del contexto.
- Estructura y coherencia: el resumen debe presentar una estructura sólida y definida y las frases deben ser coherentes, comprensibles y no contradictorias.

4.3.2.1.2. Criterios a tener en cuenta para la evaluación del contenido (*Content evaluation*)

Co-selección (*Co-selection*)

Basada en la vinculación de términos según varios tipos de similitudes.

Precisión, Re-llamada (*Recall*) y Resultados F (*F-Scores*)

Muchas de las técnicas del RTA son “sistemas supervisados de clasificación binaria”. Por “supervisión” entendemos que el programa informático necesita una “formación” o aprendizaje que deberá realizar una persona o grupo de personas. Por “binario” entendemos que habrá solo dos categorías. Por ejemplo, podrían ser criterios de clasificación la separación entre comentarios “subjetivos” y “no subjetivos” o diferenciar las frases entre “informativas” y “no informativas”.

En estos casos, podemos evaluar el clasificador utilizando *Precisión*, *Recall* y *F-score*. *Precisión* y *Recall* se calculan de la siguiente manera:

- $Precisión = TP / (TP + FP)$
- $Recall = TP / (TP + FN)$

Donde: TP significa: verdaderos positivos (correctamente clasificados como positivos). FP

significa: falsos positivos (correctamente clasificados como positivos). FN significa: falsos negativos (incorrectamente clasificados como negativos).

Un clasificador perfecto tendría tanto la *Precisión* como la *Recall* igual a 1 mientras que FP y FN serían iguales a 0.

El *F-score* es una combinación sencilla de *Precisión* y *Recall* donde el resultado perfecto sería aquel en el que el *F-score* fuera igual a 1: (Carenini, 2011:28)^[275]

$$F\text{-score} = (B^2 + 1)PR / B2P + R$$

ROC Curvas⁵⁸

Los ROC son clasificadores binarios probabilísticos. Se caracterizan por asignar a cada instancia de datos (por ejemplo, una oración) una probabilidad de pertenecer a una determinada clase de oración partiendo de las evidencias que se pueden extraer de las características que contiene.

En el cálculo de *Precisión*, *Recall* y *F-score* de un clasificador probabilístico se evalúa el clasificador con un umbral de probabilidad específico. Por ejemplo, consideramos un caso en el que la probabilidad será “positiva” si es “mayor o igual” a un umbral que hemos fijado y “negativa” en caso contrario. Un umbral comúnmente utilizado es 0,5, al ser el punto medio del rango de probabilidad [0, 1]. Efectivamente, el rango de análisis del clasificador de todas las posibilidades de probabilidad estará entre 0 y 1.

Hay una ecuación que nos ofrece la medición de las tasas de verdadero(positivo) y falso (negativo) a partir de las variaciones del umbral.

⁵⁸ ROC: acrónimo de *Receiver Operating Characteristic* o Característica Operativa del Receptor. Representación gráfica para un sistema clasificador binario cuya área posee un valor comprendido entre 0,5 y 1 y en el que 1 representa un diagnóstico perfecto. (Swets, 1996)^[324]

El cálculo de la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) se obtiene con la siguiente ecuación:

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{FPR} = \text{FP} / (\text{FP} + \text{TN})$$

Si ponemos el umbral de probabilidad del clasificador a 0, todas las lecturas con resultado igual o superior a 0 se consideran positivas, la TPR igual a 1 y la FPR igual a 0. La razón de este resultado se explica porque nada se considera negativo, con lo que $\text{FN} = \text{TN} = 0$. (Carenini, 2011:31)^[275]

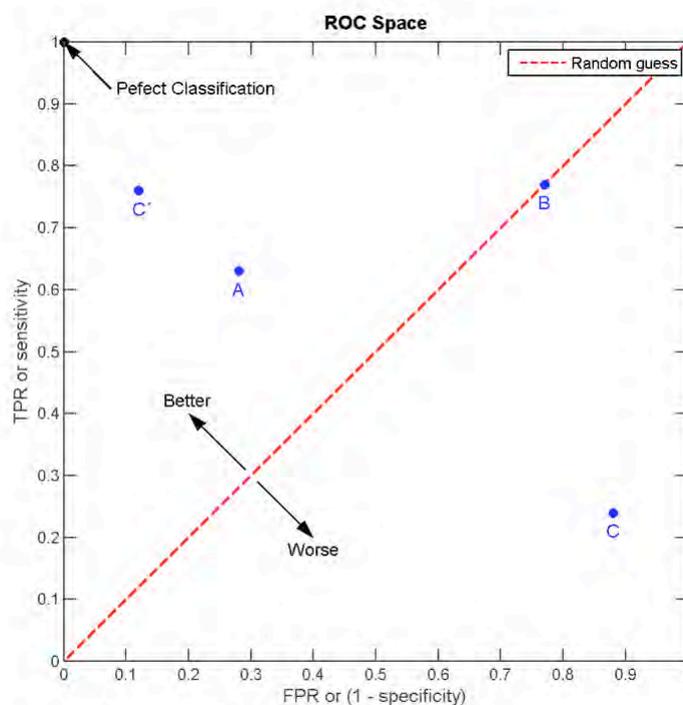


Figura 39: El espacio ROC y las parcelas de los cuatro ejemplos de predicción A, B, C y C'.

(Wikipedia, 2014)^[325]

Utilidad relativa (*Relative Utility*)

Uno de los principales problemas de los resúmenes *Precisión* y *Recall* es que los supervisores, a menudo, no están de acuerdo con la selección de las oraciones más importantes. Si se utilizan las técnicas de evaluación *Precisión* y *Recall* puede suceder que dos RTA de calidad sean evaluados de manera muy diferente. Para intentar corregir esta problemática se desarrolló la *Utilidad Relativa* (UR). Con esta técnica, cada frase obtiene un valor, llamado “utilidad de la frase”. De este modo, todas las frases pueden ser incluidas en el resumen en función de cómo sea el criterio del evaluador, el documento original y la longitud del resumen. (Ding, 2005)^[326]

Según el contenido (*Content-Based*)

Las medidas de evaluación *Co-selection* analizadas en el apartado anterior se basan en la coincidencia exacta de ciertos contenidos dentro de las frases. Este punto ignora el hecho que dos frases pueden tener la misma información aunque estén escritas de forma diferente, planteando un problema similar al de la holonimia.

Ilustramos el problema con las siguientes oraciones:

Frase 1: “*La visita del Presidente de Francia a Alemania*”

Frase 2: “*El presidente francés visita Alemania*”

Ambas frases significan lo mismo y los sistemas de evaluación *Content-based* nos permitirán identificar su similitud a partir de la identificación del contenido. (Steinberger, 2009:1008)^[297]

ROUGE (*N-gram matching*)

El sistema ROUGE es el primero de los sistemas de *Content-based* que analizaremos. Ha sido una de las herramientas de evaluación de RTA más populares de los últimos años y procede de una herramienta anterior llamada BLEU (Carenini, 2011)^[275]. El *software*

(ROUGE) se basa en el PLN y utiliza métricas que comparan un resumen (o traducción) generado de forma automática con resúmenes o traducciones realizados por personas que son tenidos como referencia.

En un inicio, la herramienta de evaluación y traducción automática BLEU buscaba la precisión entre las traducciones automáticas y las múltiples traducciones hechas por personas a través de la comparación y la superposición de n-gramas. Los n-gramas son secuencias de n palabras, de modo que todos los unigramas de un documento serían todas las palabras individuales que contiene en su interior. Análogamente, todos los bigramas de un documento serían todas las secuencias de dos palabras, y así sucesivamente.

Pyramids

Por estar fuera de nuestro uso en la investigación, apuntaremos brevemente que el método de la Pirámide es un sistema semi-automático de evaluación que identifica unidades semánticas de resumen basándose en la comparación de diversos resúmenes. (Nenkova, 2004)^[327]

Análisis Semántico Latente (*Latent Semantic Analysis, LSA*)

El LSA es un método utilizado para la extracción y representación del uso contextual del significado de las palabras a través de programas estadísticos que trabajan en “corpus” de textos de grandes dimensiones. Esta técnica reposa sobre la certeza de que la suma de todos los contextos en los que aparece (o no aparece) una palabra proporcionará un conjunto de relaciones que determinarán la similitud de significado entre palabras o entre grupos de palabras. Se utiliza regularmente dentro del PNL para la recuperación de información o en la segmentación de textos y también para la extracción de las “dimensiones escondidas del significado semántico de documentos, frases y oraciones”. Con “dimensiones desconocidas”, el investigador se refiere a los juegos de palabras, el uso de la ironía y otras expresiones cuyo significado va más allá de lo connotado en el propio texto. (Landauer, 1998:259)^[328] (Ding, 2005)^[326]

En la prensa se realizan los RAT usando el método LSA para identificar los temas importantes sin necesidad de utilizar otros recursos como WordNet. (Gong, 2002)^[329]

Es una técnica utilizada en el PLN para la extracción de las dimensiones escondidas del significado semántico de documentos, frases y oraciones, en función de su contexto. Ha sido ampliamente usada por diversas aplicaciones del procesamiento del lenguaje natural incluyendo la recuperación de información o segmentación de textos. (Ding, 2005)^[326]

4.3.2.2. Evaluación Extrínseca

Los sistemas de evaluación extrínseca evalúan RAT basados en una tarea específica y se definen como *Task-based measures*. Pueden hacerse sobre textos originales completos, resúmenes producidos por personas u otros RAT. La técnica extrínseca se inicia con una serie de preguntas a un grupo de personas respecto al resumen de forma automática que acaban de leer. Sus respuestas se compararán con las respuestas que las mismas personas dan respecto a una serie de cuestiones que hacen referencia al documento original. Mediante la comparación de ambas respuestas se determinará la calidad de los resúmenes.

Por otro lado, las métricas extrínsecas pueden evaluar la utilidad del resumen en relación a su capacidad por sustituir una tarea del mundo real.

Así pues, se considerarán mejores predicciones aquellas que obtengan una correlación mayor en relación a una tarea del mundo real. Dado que el objetivo de esta labor de resumen es de facilitar una tarea, debe ser evaluado en relación a la misma. En cuanto al desarrollo de las tareas, aunque inicialmente el sistema informático deba aprender de las acciones y criterios de una persona, si el proceso avanza exitosamente la máquina acabará por poder resolver por sí misma las tareas con igual o mayor eficiencia que las personas.

Tal y como dice Spärck-Jones: “es imposible evaluar los resúmenes correctamente sin saber para lo que son”. (Spärck-Jones, 1998:10)^[277]

4.3.2.2.1. Sistemas de evaluación extrínseca

Según la categorización de documentos (*Document Categorization*)

Basado en asignar categorías o clases a los documentos que quieren categorizarse según su característica en cuanto a tipo de texto (autor, año de edición, etc.).

Recuperación de la información (*Information Retrieval*)

La recuperación de la información puede servir como sistemas de evaluación de la calidad de un resumen.

Según los criterios de relevancia o pertinencia

El *Relevance Assessment Task* es uno de los sistemas de evaluación extrínseca más utilizados. La “tarea” consiste en asignarle a un sujeto un tema y un documento y pedirle que determine la relevancia de dicho tema en relación al documento. Posteriormente, en el RAT, se analizará la influencia del tema en el resumen y se valorará la precisión en la valoración de la *relevancia* junto con el tiempo empleado que ha tardado el sujeto en realizar la tarea. Un resumen cuyo tema coincide con el que el sujeto ha señalado y que, además, para señalarlo el sujeto ha tardado un tiempo elevado se considerará satisfactorio. (Mani, 2001:241)^[299]

Capacidad de respuesta (*Question answering*)

Este sistema evalúa la cantidad de respuestas correctas realizadas por una serie de sujetos ante la lectura de diversos documentos: los documentos originales, un resumen generado de forma automática o un resumen generado por un profesional. Los sujetos tienen que responder a las preguntas y se compararán sus respuestas. (Steinberger, 2009:1012)^[297]

Comprensión de lectura (*Reading comprehension*)

En consonancia con el *Question answering*, el sistema de evaluación de la calidad de un RAT a partir del *Reading comprehension* consiste en realizar una serie de cuestiones a un sujeto acerca del documento que ha podido leer. Se podrá valorar la comprensión del documento si las respuestas del texto completo son similares a las respuestas del RAT. En esta serie de casos se considera que el RAT puede sustituir al texto completo. (Morris, 1999)^[330]

Con el amplio desglose y estudio que acabamos de presentar teníamos como objetivo mostrar la gran variedad de posibilidades que han surgido en la materia. Este perfeccionamiento metodológico continúa dándose actualmente y lo explicamos tanto por la mejora de las capacidades de las computadoras como por la gran utilidad de los resultados obtenidos con dichos métodos. En capítulos posteriores mostraremos aquellos que hemos elegido para nuestra investigación, así como los motivos principales que explican el por qué de su elección.

4.3.3. Detección de la subjetividad y los sentimientos

Hemos incorporado deliberadamente la aplicación del RAT en el estudio de los sentimientos dentro del capítulo dedicado a la metodología. Con ello queremos dar a entender que es posible detectar y extraer de forma metodológica la expresión de sentimientos en diversos textos usando el RAT y sus técnicas de una determinada manera. Dedicaremos el siguiente apartado (2.5) de este segundo capítulo a desarrollar dicho método con mayor detalle, bajo el término que hemos definido anteriormente: *Sentiment Analysis*. Nuestro objetivo en el punto presente es explorar cómo podemos localizar y estudiar la expresión de la subjetividad dentro de un texto, es decir, en cómo logramos dilucidar que están mostrando su opinión acerca de algo y no, por ejemplo, una descripción sin introducir valoraciones. El paso siguiente, realizado casi en paralelo, consistirá en detectar e identificar los sentimientos. Como referencia inicial, avanzaremos en este punto de la mano de los trabajos académicos y de investigación que se han llevado a cabo en esta materia.

La investigación relacionada con la identificación de las opiniones en el lenguaje subjetivo utiliza un amplio abanico de nombres para referirse a la misma cosa y que utilizaremos de forma indistinta: detección de opinión, detección de la subjetividad, clasificación de la polaridad u orientación semántica, entre otros (Pang, 2008:36)^[331]. El concepto que nos permite ampliar la información acerca de los sentimientos que vayamos encontrando es el que denominaremos “granularidad” y puede llegarnos a indicar, por ejemplo, qué aspectos de un producto o servicio le gustan o no le gustan a un usuario particular y el por qué. Según los niveles de granularidad podremos también detectar cuándo los usuarios manifiestan su acuerdo o desacuerdo o discuten acerca de algún tema en particular. (Carenini, 2011)^[275]

Situándonos en un nivel muy detallado podríamos llegarnos a plantear cuestiones como: “¿A la gente le gusta este candidato?” o “¿Qué piensan los consumidores de este producto?”. A mayor nivel de granularidad, mayor ímpetu en detallar lo que a un determinado consumidor le gusta o le disgusta de un producto y por qué razones.

A continuación describiremos, de menor a mayor granularidad, varias tareas relacionadas con la detección de la subjetividad. (Carenini, 2011:55)^[275]

4.3.3.1. Clasificación del sentimiento en Internet

Las empresas, al igual que los partidos políticos, están muy interesadas en identificar lo que la gente piensa de ellas, ya sea de su producto o de sus líderes. De las diversas formas de acceder a ese conocimiento, una opción viable es la de buscar comentarios y opiniones de un grupo “test” creado a pequeña escala y formado por *influencers*. Si la ambición de la empresa es obtener las opiniones de miles o millones de personas, puede analizar los comentarios que ese volumen de usuarios realizan en internet (comentarios en periódicos digitales, comentarios en sitios web, blogs o foros, entre otros). En función de las necesidades de la empresa, el partido político o la institución, el nivel de granularidad podrá oscilar significativamente, aunque en la mayoría de las ocasiones se obtendrá simplemente una indicación o aproximación a tendencias favorables o desfavorables o si la polaridad es de índole positiva o negativa. (Carenini, 2011:56)^[275]

4.3.3.2. Clasificación del sentimiento de productos

En otras ocasiones, lo que queremos saber son las valoraciones que puede haber recibido un producto en concreto. Es decir, queremos saber si “ha gustado o no” y para ello precisaremos saber su polaridad (positiva o negativa) y el nivel (poco o mucho). Imaginemos el análisis de una película: habrá quienes la hayan encontrado nefasta y quienes sencillamente se han aburrido viéndola. En el primer caso, sería nos situaríamos en una polaridad negativa de un nivel más acusado (película nefasta) que en el segundo caso (película aburrida). El estudio, sin embargo, se podría complicar: a un espectador le puede haber gustado la película pero no la actuación de los actores principales. En este caso, se complica también el resultado del estudio de los sentimientos. ¿Deberían ser positivos, negativos o neutros?

Desde Amazon a la tienda de productos artesanales Etsy, son multitud las webs que ofrecen a sus usuarios la calificación de sus productos (votos en forma de corazones, estrellas, etc.) según si han quedado satisfechos o no con la compra. Aunque presentan una simplificación, pueden servirnos de orientación y, más allá de eso, constituirá una las partes fundamentales para el análisis de los sentimientos (al mismo tiempo, suponen una información que los próximos clientes consideran sólida y fiable)^[332]. (Cui,2006)^[333]

4.3.3.3. Clasificación del sentimiento en frases (textos)

La transcripción de una reunión donde las personas opinan sobre un producto permite, de forma rápida y económica, obtener las valoraciones que se han hecho durante el encuentro. Para ello requeriremos el uso de una herramienta (*software* o algoritmo) preparada para la identificación y clasificación de sentimientos. Si lo necesitamos, podremos preparar la misma herramienta de análisis para que seleccione solo las informaciones o frases más relevantes y su polaridad (positiva o negativa). (Kim, 2004)^[334]

4.3.3.4. Clasificación de Facetas del Sentimiento

Si la voluntad es profundizar más allá de un “me gusta” o “no me gusta” del consumidor y queremos encontrar las razones que tiene para elegir o rechazar un producto (qué sentimientos le despierta o qué opinión le merece), podemos investigar sobre otras facetas o aspectos del producto y que han contribuido significativamente a la opinión final. (Liu, 2012)^[335]

Como ocurre en muchas de las investigaciones en el campo del *Text Mining* y del PLN, existen enfoques supervisados y no supervisados. La mayoría de las técnicas de clasificación mencionadas anteriormente pueden ser “supervisadas” de tal modo que sea el propio algoritmo el que reciba la “formación” para poder diferenciar los datos positivos de los negativos en cada oración recogida dentro de un documento. Los algoritmos, con o sin supervisión, pueden predecir la polaridad a partir de unos datos etiquetados. El enfoque semi-supervisado permitirá identificar de forma manual una serie de palabras con una determinada polaridad que servirán de base para el etiquetaje y la clasificación automática posteriores. La ejecución de este proceso puede conducir al descubrimiento de nuevas palabras, a la incorporación de nuevas palabras clave (de base) y también a la repetición del proceso. Aquí es importante retomar la mención acerca de la divergencias entre los enfoques basados en el léxico y los enfoques estadísticos. Un enfoque basado en el léxico partirá de un diccionario de palabras generalmente asociadas con calificaciones numéricas (fuerza de la polaridad de cada palabra). Por ejemplo, los resultados podrían oscilar entre +/-5, siendo +5 la calificación de un sentimiento muy positivo (“fantástico”) y siendo -5 la calificación de un sentimiento muy negativo (“terrible”). A partir de un texto, el sistema basado en el léxico identificaría las palabras que figuran en su diccionario y asociaría a cada una de ellas las diversas calificaciones o puntuaciones que les debe corresponder. En una versión simplificada, un texto se constituiría de la operación entre las polaridades de las distintas palabras analizadas. (Carenini, 2011:57)^[275]

4.3.3.5. SOCal (*Semantic Orientation Calculator*)

El SOCal, como sistema basado en el léxico, es notablemente más sofisticado porque no considera suficiente la simple suma de las puntuaciones de las palabras. Según queda demostrado, el análisis semántico de la totalidad de la oración permite obtener una puntuación más adecuada que partiendo de la valoración segmentada en cada una de sus palabras.

En referencia a los sistemas estadísticos que se aplican en la actualidad, muchos de ellos no se basan en diccionarios hechos a mano sino en diccionarios que se desarrollan automáticamente a base de la incorporación de términos y frases subjetivas obtenidas de datos etiquetados previamente (o, incluso, sin etiquetar). Una técnica posible para ello es la creación de una lista de palabras subjetiva gracias a la identificación de las palabras que se presentan con mayor frecuencia en el texto etiquetado. Otros sistemas estadísticos no utilizan listas de palabras sino que extraen los rasgos léxicos primarios como *unigrams* o *bigrams*; dejan que el método de aprendizaje de la máquina continúe de forma automática y como resultado se produce una correlación de las características de significado positivo o negativo (Taboada, 2011:270)^[336]

4.3.4. Uso de grafos en los resúmenes automáticos de textos

A lo largo del capítulo dedicado a los antecedentes del grafo, vimos como pueden representar una gran cantidad de objetos, estados o hechos: desde personas conectadas a través de redes sociales a páginas conectadas por hipervínculos o estaciones de servicio conectadas a través de rutas, las posibilidades de representación del grafo son, sin duda, uno de sus virtudes destacadas. En este sentido, la ventaja de la Teoría de grafos es que, una vez que una situación o aspecto se ha representado correctamente, las técnicas pueden aplicarse sin problema para su análisis. (Sizov, 2010)^[303]

Los RAT también se pueden representar mediante grafos y, en consecuencia, se pueden analizar a partir de las técnicas que permite la teoría.

Hay una multiplicidad de opciones para representar un texto a través de un grafo. Los nodos representarán las palabras, frases o párrafos. Si los nodos quedan convertidos en texto, los enlaces se convertirán en:

- Relaciones de semejanza
- Relaciones semánticas

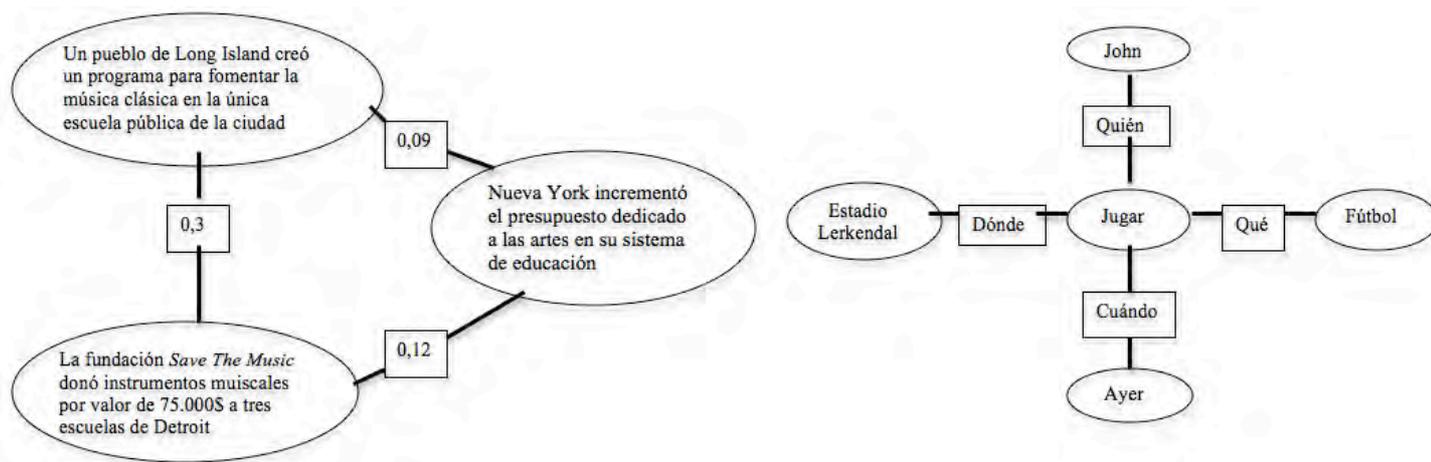


Figura 40: En la figura izquierda encontramos un grafo basado en las relaciones semánticas (según el significado de las palabras y las oraciones). En la figura derecha encontramos un grafo basado en las relaciones de semejanza (términos vinculados según una relación concreta entre ellos). Los grafos semánticos capturan información con mayor precisión y son más útiles desde el punto de vista teórico. (Sizov, 2010:38)^[303]

4.4. Aplicaciones del Resumen Automático de Textos

La aplicación del RAT está siendo esencial para que nuestras sociedades puedan seguir manejando, comprendiendo y organizando el vasto volumen de información generado con los nuevos medios de comunicación y la evolución de la red. La necesidad de establecer relaciones entre diversos hechos o sucesos dispersos, ya sean temáticas o causales, para intentar obtener datos relevantes a favor de la correcta toma de decisiones futuras (recordamos aquí la trascendencia de planteamientos como el *Small World*) sigue siendo fundamental pero, sin la ayuda de los RAT, sería imposible. Como hemos indicado en la sección de métodos, las aplicaciones entran dentro del campo del PLN. Algunas de las aplicaciones más comunes ya las hemos señalado, como la corrección de textos, la síntesis de discursos o el reconocimiento de voz. Otras aplicaciones habituales pueden ser la generación de noticias con un sistema como el de Google News. (Alami, 2015)^[337]

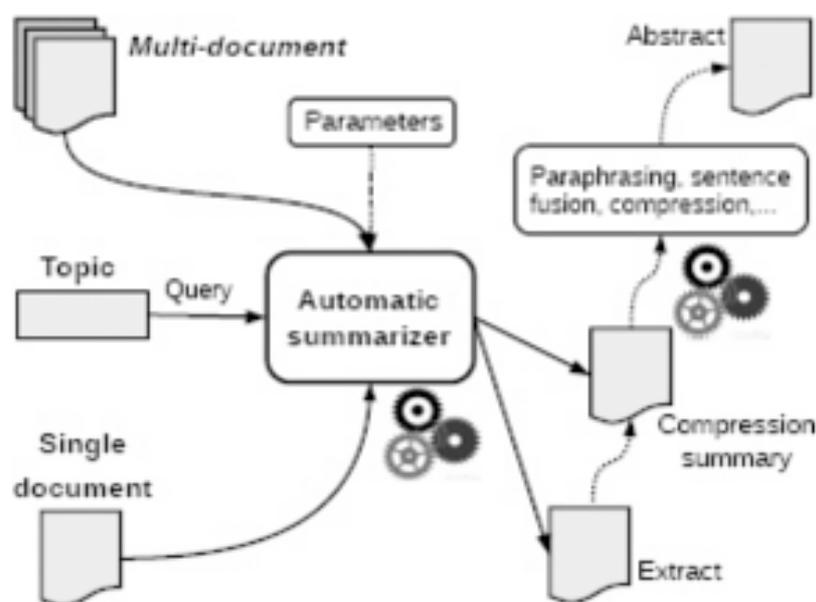


Figura 41: Proceso simplificado de RAT y elaboración de un resumen de salida. (Torres-Moreno, 2014:14)^[280]

Columbia Newsblaster
Summarizing all the news on the Web

Thursday, August 6, 2015
Articles from 08/03/2015 to 08/06/2015
Last update: 2:30 PM EST

Search for:
Offline summarization
Go

U.S.
World
Finance
Sports

View Today's Images
View Archive
About Newsblaster
About today's run
Newsblaster in Press
Academic Papers

Article Sources:
reuters.com (37 articles)
abcnews.go.com (23 articles)
latimes.com (23 articles)
cbsnews.com (22 articles)
dallasnews.com (21 articles)
foxnews.com (9 articles)
harrisburg.com (8 articles)
baltimoresun.com (4 articles)

Eclipsed by Trump, other Republican candidates vie for some spotlight at debate (U.S., 11 articles) [UPDATE]
Donald Trump will be at center stage and potentially in the line of fire on Thursday as 10 leading Republican presidential candidates look for a break-out moment at their first debate on the road to the November 2016 election. Flanking Trump in the 10-candidate debate Thursday will be Florida Gov. Jeb Bush, the former governor of Florida and onetime GOP front-runner, and Wisconsin Gov. Scott Walker Fox, the broadcast sponsor for Thursday's session, announced in the spring that 10 candidates would get to debate, picked based on who has the highest average standings in the five most recent national polls released by Tuesday evening. The debate is expected to be dominated by the billionaire celebrity candidate whose rise to the top of the polls has tapped populist discontent and sent shock waves through the Republican establishment. The prime-time spotlight will shine on 10 candidates, leaving Perry and six other GOP laggards to a consolation debate four hours earlier. The GOP debate drama is growing Thursday is the big Republican presidential debate there are seventeen candidates but only the top ten. Karen Travers Karan and debate will be on fox news with seventeen candidates vying for the White House how is fox going to determine the top ten.

Other stories about Trump, candidates and debate:

- Democratic primary voters see Hillary Clinton as most electable (4 articles) [UPDATE]
- Ruben Navarrette: Trump might be helping immigration debate (4 articles) [UPDATE]

Top News

Taylor Swift Says Her Friendship With Kanye West Was 'Important' to Jay Z Video (U.S., 16 articles)
Now Playing: Taylor Swift Says Her Friendship With Kanye West Was Important to Jay Z Now Playing: Kermit and Missy Piggy Call It Quits, Taylor Swift Explains Kanye Friendship and More in Pop News. Now Playing: Kentucky Deputy Sheriff Handcuffs Disabled Boy; Pres. Obama Hosts White House Demo Day; Giant Minion Blocks Traffic in Ireland. Now Playing: The Truth About Fertility: How Much Is a Human Egg Worth? Now Playing: One Woman's Journey to Find Her Birth Mother: Part 1.

Plane window, other debris found on Reunion Island, Malaysia transport minister says (U.S., 5 articles) [UPDATE]
A Malaysian recovery team on the French island of Reunion has collected an aircraft window and other plane parts but could not immediately confirm whether they belonged to Malaysian Airlines Flight 370, Transportation Minister Liow Tiong Lai said Thursday. Malaysian Prime Minister Najib Razak said early Thursday that the wing part, known as a flap, was confirmed to have come from Flight 370, which disappeared without a trace 17 months ago with 239 people aboard. Shim's wife, Christine Tan, was the lead flight attendant aboard the flight that vanished after taking off from Kuala Lumpur last March.

How "Daily Show" host Jon Stewart became a star (U.S., 6 articles) [UPDATE]
NEW YORK His "Daily Show" set is being donated to Washington's Newseum and 19 of his custom-made show suits have been auctioned off for a disability charity. Stewart, who announced his departure in February, leaves thousands of grieving fans and an outpouring of media tributes debating his legacy as the nation's most popular political satirist. The comedian had just become a mega-star thanks to his show's satirical coverage of the 2000 presidential election, which Stewart and his team aptly dubbed "Indecision 2000."

Obama presses critics on Iran nuclear deal (World, 9 articles) [UPDATE]
LONDON The first international container ships began arriving in Iran this week after the nuclear deal between Tehran and world powers, yet many ship owners remain wary of resuming business until sanctions are removed - still some months away. WASHINGTON Obama administration officials told lawmakers on Wednesday that sanctions relief under the Iran nuclear deal will not be allowed to be used to bankroll terrorism and said the sanctions can be "snapped back" quickly if Tehran violates the agreement. The White House is conducting a lobbying blitz to convince Congress to back the agreement announced on July 14 between the United States, five other world powers and Iran.

Figura 42: Captura de pantalla del servicio de selección de noticias Newsblaster. (Newsblaster, 2015)^[338]

Como ejemplo sumamente exitoso (especialmente en la década anterior) a nivel de usuario recordamos el servicio RSS (*Rich Site Summary*⁵⁹). Este tipo de *feed summarization* realiza resúmenes de diversas fuentes *online* (blogs, webs, etc.) y los presenta en el propio ordenador. Versiones más modernas del servicio RSS y que parten de su misma filosofía pueden ser *Feedly*⁶⁰ o *Flipboard*⁶¹. (Hu, 2007)^[339]

⁵⁹ Los RSS (*Rich Site Summary*) han evolucionado a *Really Simple Syndication* cuya utilidad reside en poder compartir contenido de sitios web e información actualizada. (Wikipedia, 2015)^[343]

⁶⁰ www.feedly.com

⁶¹ www.flipboard.com

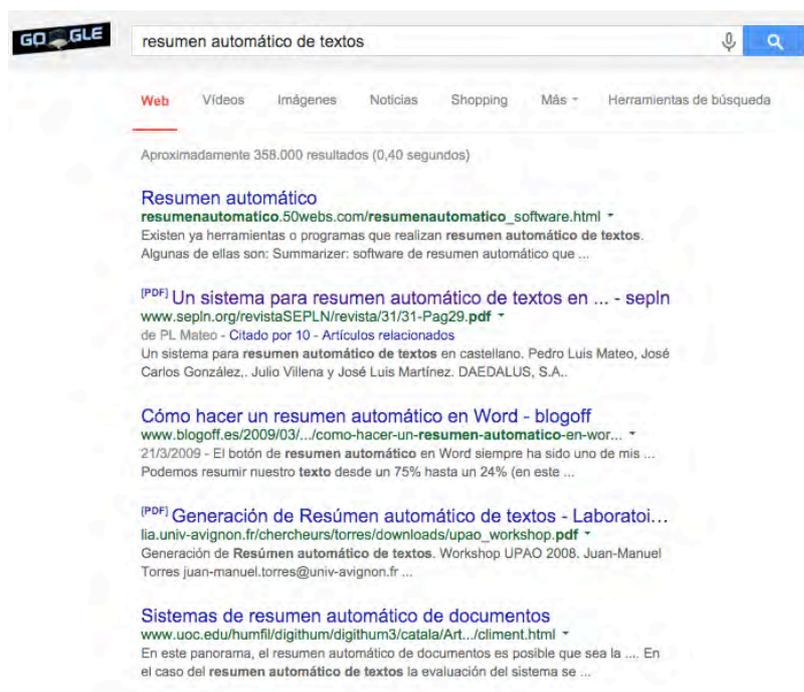


Figura 43: Captura de pantalla de un resultado de búsqueda en Google. Fuente: (Google, 2014) [consulta: 13-11-2014 16:08]^[344]

Encontraremos también herramientas para el resumen de tuits (Chakrabarti, 2011)^[340], el resumen de sitios web (ver figura 43) (Sun, 2005)^[341] (Buyukkokten, 2001)^[342], correos electrónicos (Rambow, 2004)^[345], (Carenini, 2011)^[275], informes empresariales o políticos, resúmenes de reuniones o discursos, opiniones, biografías (Biadsky, 2008)^[346], el RAT (Banko, 2000)^[347] y también en dominios específicos: medicina, química, derecho, entre multitud de campos del conocimiento. (Yousfi-Monod, 2010)^[348] (Liu, 2009)^[349]

Retomando nuestro objetivo a la hora de realizar el trabajo de campo en Twitter, nos centraremos en el estudio de las investigaciones desarrolladas en los RAT de los medios sociales y de servicios. Dado que el trabajo de campo de esta tesis se ha centrado en el medio social Twitter, consideramos oportuno centrarnos en las investigaciones desarrolladas sobre los resúmenes de textos de los medios sociales y de otros servicios relacionados con internet (correo electrónico o blogs).

El crecimiento exponencial del uso de los medios sociales, con Facebook⁶² y Twitter⁶³ a la cabeza, les sitúa como actuales abanderados de la masificación de las nuevas tecnologías y del uso de internet. Sabemos que el RAT cuenta con más de cincuenta años de desarrollo y que las aportaciones a su mejora son constantes. Sin embargo, con el surgimiento del *Big Data* necesitamos todavía más y mejores sistemas con los que trabajar la información.

Si decidimos, dentro de las empresas de servicios de los medios sociales, analizar los textos compartidos por los usuarios de Twitter, obtendremos un volumen inmenso de información pero desestructurado y heterogéneo que, a la práctica, resultará inmanejable. Precisaremos algún tipo de filtro para evitar tuits sin texto o demasiado breves, retuits, palabras abreviadas o emoticonos. Un medio de dichas características dificulta los análisis tradicionales ya que fueron ideados para funcionar con textos más amplios y estáticos.

Realizaremos a continuación una breve panorámica de los resúmenes más habituales en el entorno de internet: correos electrónicos, blogs y chats, Facebook y Twitter. Será con una breve visión de cada uno de ellos que podremos apreciar los cambios que requiere el tratamiento de la información en un entorno puramente digital. Ello nos dará el acceso a entrar cada vez con mayor concreción en el análisis de Twitter y nos permitirá situarnos a nivel práctico y enfocar los objetivos de nuestra investigación.

⁶² Facebook pasó de tener 100 millones de usuarios en 2008 a 1.441 millones de usuarios en abril de 2015. <Statista, 2015. En línea>

⁶³ Twitter ha pasado de 151 millones de usuarios en el 2º trimestre de 2012 a 302 millones de usuarios en el 1º trimestre de 2015. <KIM, 2015. En línea>

4.4.1. Resumen de correos electrónicos

El tráfico de los buzones de correo electrónico es constante y algunos de los mensajes se extienden en conversaciones que pueden durar días, semanas o meses. Además, el tiempo que dedican los usuarios a revisar la bandeja de entrada suele ser insuficiente para hacer una lectura completa de todos los correos. En un uso cotidiano como el que describe Muresan, disponer de un breve resumen de la información más importante del correo ayuda a decidir de antemano la conveniencia de su lectura y permite ahorrar tiempo al decidir si merece una pronta respuesta o es mejor eliminarlo. Siguiendo un sistema de clasificación, podemos distinguir entre los correos individuales y las conversaciones entre varios usuarios. (Muresan, 2001)^[350]

El resumen de los correos electrónicos individuales es muy útil para su selección y visualización y es especialmente práctico si se consultan desde un dispositivos móviles (cuyo espacio es muy limitado). Si queremos extraer un resumen de las conversaciones de los correos electrónicos deberemos seguir un proceso diferente al de otros tipos de comunicaciones escritas. Al usar una estructura de diálogo inherente propia necesitaremos un sistema de resumen extractivo basado en el *Machine Learning*. (Rambow, 2004)^[345]

Varios investigadores han propuesto aproximaciones distintas a los resúmenes de textos de correos electrónicos. Por un lado, se ha puesto el centro en las discusiones desarrolladas sobre grupos de noticias (Newman, 2003)^[267]. Por otro lado, el centro de la investigación se ha focalizado en el análisis de las discusiones y se ha tenido en cuenta que la conversación representa un proceso en la toma de decisiones (Wan, 2004)^[351]. Desde otro punto de vista, se ha investigado la elaboración de resúmenes indicativos sobre la identificación del tema central de la conversación con el objetivo de lograr ofrecer la información más relevante y que permita al usuario decidir si desea o no recuperar el hilo de la conversación (Nenkova, 2003)^[347]. Con el RAT aplicado a los correos electrónicos y con el uso de diversos métodos de análisis pueden llegarse a identificar variables básicas del usuario (como su género y edad) a través de las características léxicas de su discurso, la longitud de las frases, los patrones de

participación en las conversaciones o el uso de determinadas composiciones gramaticales. (Garera, 2009)^[352]

Todos los resúmenes empiezan a partir del texto del asunto. Luego, se selecciona una frase central del contenido del correo a partir de la superposición de términos relacionados con el mismo contenido expresado en el asunto. El resto de las oraciones se seleccionarán según el nivel de superposición de términos relacionados con la frase central. El objetivo será poder describir sintéticamente el contenido básico del correo electrónico o identificar el tema en concreto del que trata. (Nenkova, 2003)^[347]

Carenini et al. creó un sistema de resumen de correos electrónicos sin necesidad de supervisión construido a partir de palabras clave que hacen de referencia. Su enfoque se basa en analizar la estructura de las conversaciones mantenidas y las palabras que se repiten a lo largo de los correos. El sistema se representa con grafos, donde el correo electrónico se fragmenta en nodos y los nodos pueden representar frases, fragmentos o participantes. (Carenini, 2007)^[353]

Un ejemplo de la aplicación de los resúmenes de correos electrónicos lo encontramos en su uso para procesos de investigación criminal o forense. La mayoría de los países ya aceptan los correos electrónicos como pruebas válidas para utilizar en juicios (Gupta, 2004)^[354]. Entre multitud de ejemplos, en Estados Unidos se han utilizado en el juicio antimonopolio contra Microsoft o en la investigación sobre el escándalo de Enron. Para hallar un correo o una sentencia que sea válida como prueba en un juicio (por ejemplo), el volumen de datos que es preciso analizar es enorme. Un buen método para extraer y resumir aquellas conversaciones más relevantes es realmente valioso y puede permitir ahorrar un tiempo indispensable en el desarrollo del juicio. Las herramientas capaces de seleccionar y discernir el contenido relevante en función del ámbito de investigación serán las aliadas imprescindibles para los cuerpos de seguridad, policía o Hacienda, entre otros. (Carenini, 2011:8)^[275]

4.4.2. Resumen de Blogs y Chats

Los blogs, tanto los personales como los de carácter profesional, son ya un pilar fundamental en internet y en según qué casos (blogs de celebridades, políticos e *influencers*) han entrado como parte relevante en la elaboración y difusión de la información de actualidad. Inicialmente, el blog se mantenía con un dominio separado en la red, aunque la tendencia actual es a centralizar el tráfico e incorporar el blog dentro del sitio web. Como apuntó en su momento Ku, a la gente le gusta expresar sus opiniones y les gusta conocer la opinión de los demás. A medida que la gente los empezó a mostrar cada vez más en los comentarios en weblogs (o blogs) la investigación en la blogosfera ganó progresivamente en popularidad e importancia. (Ku, 2006)^[355]

El blog presenta un dinamismo superior al de la web y se caracteriza por ofrecer una gran cantidad de entradas que se ubicarán en diferentes categorías definidas previamente por el autor. Las entradas podrán organizarse automáticamente por orden cronológico y por categorías. Las entradas de un blog pueden tratar de todo tipo de tema, y la tendencia actual está llevando a su especialización (blog de tendencias urbanas, política económica europea, cine asiático, etc.); sus autores los definen por en función de muchas variables, desde los intereses a las aficiones o a la profesión. En según qué casos, los contenidos pueden resultar de sumo interés tanto para lectores individuales como para empresas con pretensiones comerciales. En otros, puede que los organismos públicos obtengan información valiosa o, en otros, los beneficiados sean los investigadores de varios campos del conocimiento (psicología, sociología, antropología y, por supuesto, el marketing y la publicidad).

Las técnicas de *Data Mining* han demostrado ser óptimas para descubrir información y obtener conocimientos a través del contenido de los blogs; puede estar formado por una enorme cantidad de blogs diferentes entre sí, categorías, “*topics*” y entradas. Como es de suponer (y manejando cifras como la que apuntábamos con anterioridad y que hablaba de la creación de más de 300 blogs en el mundo cada 60 segundos), la aspiración a resumir tal cantidad de información es un verdadero reto para los académicos y estudiosos de la materia. Avanzar hacia un resumen efectivo solo será posible si el análisis cuenta con las

especificidades propias de este medio de comunicación: el tema que tratan, los intervalos de tiempo entre las entradas (*posts*), la composición del título de las entradas, la estructura de los enlaces y los comentarios. Las técnicas de RAT procedentes de blogs se centran en el resumen realizado vía abstracción y extracción. Pero, como apuntamos, la abstracción es más complicada de automatizar porque genera resúmenes que funcionan como paráfrasis de un conjunto de documentos muy amplio. Por esta razón, la mayoría de las técnicas se centran en la extracción. (Kolcz, 2001)^[356]

Existen varias aplicaciones para el resumen de textos especializados en blogs y hemos convenido en señalar las siguientes. (Asbagh, 2009)^[357]

Maximum Marginal Relevance Multi Document (MMR-MD): método de resumen extractivo basado en el concepto de *Maximal Marginal Relevance* usado para la recuperación de la información. Puede incluir diversas variables para la selección de las oraciones, tales como las palabras clave, el orden cronológico, la similitud entre *topics*, los criterios anti-redundancia y una penalización por pronombres (Carbonell, 1998)^[358]. El MEAD es otro método de resumen extractivo que opera en el nivel de frases u oraciones y que determina la jerarquía de las palabras según su frecuencia (Radev, 2000)^[283]. Finalmente, señalamos el método OCELOT: es un sistema automático de resumen de webs que utiliza modelos probabilísticos para identificar su “esencia” (equivalente al hallazgo del tema principal dentro de una conversación de correo electrónico). El interés de los investigadores por seguir desarrollando técnicas de resumen específicas para los blogs continúa a día de hoy, prueba de ello serían los métodos de detección de *spam* en los comentarios de las entradas de un blog, el etiquetaje (*tagging*) y la *Opinion Mining* (Zhou, 2006)^[359]. De todos modos, pocos estudios se han publicado al respecto y todavía hay mucha labor de investigación por desarrollar. (Asbagh, 2009)^[357]

4.4.3. Resumen de mensajes en Twitter

Hemos elegido Twitter para la investigación de esta tesis al ser una plataforma de servicios de red social que ofrece un rápido acceso a la difusión de noticias y opiniones. Cuando tiene lugar un acontecimiento relevante, los usuarios, instantáneamente, publican tuits que detallan, amplían o matizan la información sobre el acontecimiento. Debido a las constantes actualizaciones en tiempo real del servicio, se produce un flujo de información que puede viralizarse rápidamente. Hay investigadores que han demostrado que mediante una apropiada monitorización de Twitter, es posible identificar acontecimientos que están sucediendo en tiempo real (Sayyadi, 2009)^[360] (Popescu, 2010)^[361] (Watanabe, 2011)^[362] tales como terremotos (Sakaki, 2010)^[363] protestas políticas, disturbios o, en otro orden de eventos, lanzamientos de nuevos productos al mercado. (Chua, 2013)^[364]

Desde su aparición en 2006, Twitter ha crecido de una forma que es difícil de encontrar en una época anterior. Según el propio sitio web oficial de Twitter⁶⁴, las cifras de uso de su plataforma son las siguientes:

- Usuarios activos cada mes: 302.000.000
- Tuits enviados cada día : 500.000.000

Con este volumen de tráfico de usuarios y tuits, la plataforma se ha convertido en una fuente destacada para obtener información en tiempo real acerca de casi todos los temas de actualidad posibles. Se ha constatado en numerosas ocasiones que Twitter ha sido el primer medio para difundir acontecimientos extraordinarios, incluso antes de que los medios de información tradicionales lo hayan hecho. El caso paradigmático es el del aterrizaje del vuelo US Airways en el Río Hudson. (Beaumont, 2009)^[365]

⁶⁴ <https://about.twitter.com/company>. En línea 20141117. 18:30.

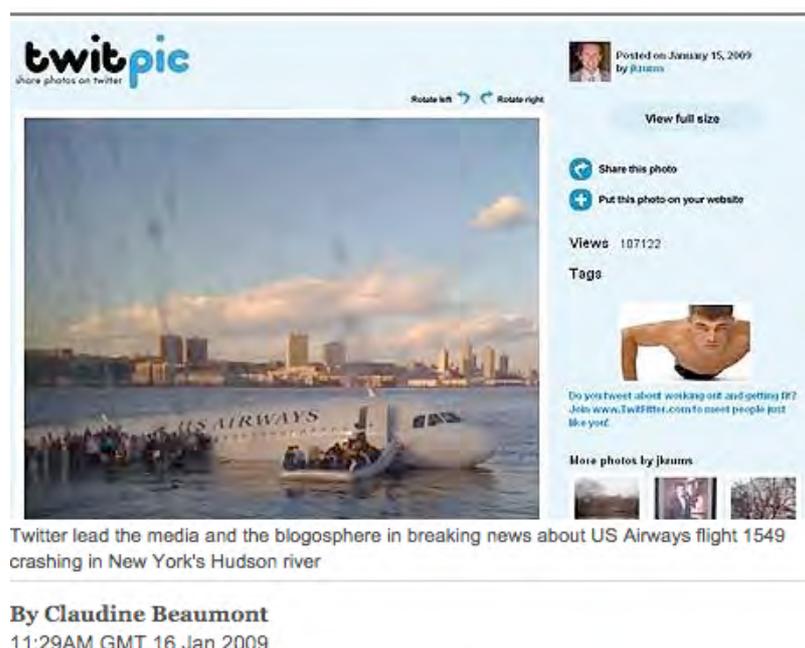


Figura 44: Imagen del accidente enviada a través de una cuenta personal de Twitter que convirtió a la plataforma en el primer medio en difundirla. (Telegraph, 2014)^[366]

En los dos primeros puntos de la tesis hemos hablado del funcionamiento de Twitter todavía de manera tangencial, apuntando simplemente algunos rasgos o características fundamentales pero centrándonos en construir el “corpus” que circunda al medio social y que permite tanto su existencia como el análisis de lo que acontece con su uso masivo. Apuntaremos ahora una doble dinámica que se observa en el uso de Twitter: una parte considerable de las conversaciones se desarrollan en pequeños grupos de amigos o compañeros de trabajo pero, en otras situaciones totalmente distintas, en las conversaciones intervienen cientos, miles o cientos de miles de personas con sus correspondientes tuits (desde un tuit a decenas). La creciente popularidad de Twitter está permitiendo que, en determinados momentos y ante determinados sucesos, una gran masa crítica de usuarios hable sobre un mismo tema que suele quedar englobado rápidamente con uno o varios #hashtags elegidos normalmente por los propios usuarios. Este ingente volumen de información puede ser resumido de forma automática, de manera que se puede obtener información de todas las comunicaciones relevantes sobre un tema y prácticamente en tiempo real. Ello significa que los periódicos y la prensa en general dejan de marcar la agenda (*agenda setting*) en función de lo que consideren importante. Los propios usuarios deciden acerca de qué tema quieren hablar o de qué forma quieren abordarlo, sea o no “reconocido” por los medios. Como es sabido, una

gran cantidad de los tuits que se realizan en momentos de actividad masiva están directamente relacionados con acontecimientos, eventos (conciertos, eventos deportivos, festivales, entregas de premios) que han dado lugar a algunas investigaciones concretas. (Chakrabarti, 2011)^[367]

En cuanto al estudio de todo este material que comparten los usuarios, ha quedado patente que existen numerosos métodos para el resumen automático de textos (desarrollados en los puntos anteriores). Sin embargo, existen poca documentación y escasos trabajos de investigación relacionados con el resumen de grandes cantidades de textos. Presentan características muy diferentes: son poco estructurados, tienen un alto nivel de repeticiones y, en general, contienen mucha información poco relevante, comúnmente denominada “ruido”. (Das, 2007)^[279]

Una de las referencias en el campo de la investigación de la comunicación en Twitter (también llamado *microblogging* de forma genérica) es Beaux Sharifi. El investigador ha presentado resúmenes que sitúan a Twitter como fuente de información y ha llamado al resultado del trabajo Resumen Automático de Microblogs (Sharifi, 2010)^[368]. Sus características principales identifican y resumen los *topics*⁶⁵. (Sharifi, 2010)^[368]

⁶⁵ Por “Topic” entendemos temas de debate o discusión e incorporamos oraciones y palabras clave compartidas por un segmento específico de usuarios.

En resumen, en este punto del capítulo hemos enfocado nuestros esfuerzos en demostrar la valía del Resumen Automático de Textos como herramienta imprescindible para analizar desde diversos ángulos el contenido de un texto y obtener del análisis conclusiones relevantes que puedan aplicarse en diferentes campos en función de las necesidades de cada investigación. Este punto guarda una estrecha relación con lo descrito en los capítulos anteriores; en el punto uno de este capítulo quisimos mostrar la diferencia entre las técnicas cuantitativas y cualitativas de investigación en el terreno de la comunicación, remarcando que para nuestro cometido precisábamos centrarnos en las técnicas cualitativas. Las técnicas cualitativas son las que se basan en la interpretación de los textos que se han realizado, ya sea con el uso de las técnicas tradicionales (entrevistas, grupos focales, etc.) o de las más novedosas, que hemos tratado en los siguientes puntos. Antes de llegar al despliegue de la Teoría de grafos, pilar que ha permitido organizar, dentro del ámbito de la investigación en comunicación (entre otros), la información en redes y estudiar las relaciones dentro de ellas, dedicamos un espacio a comentar métodos propios de investigación de *insights*. Asimismo, mostramos algunos ejemplos concretos de cómo las empresas han desarrollado, por su lado, su propia metodología. Con ello, intentamos señalar que nuestra propuesta, que utiliza los últimos avances de la Teoría de grafos aplicados al RAT, es totalmente compatible y complementaria a los métodos que las empresas siguen desarrollando por su cuenta. Nuestro análisis de un servicio de medios sociales como el de Twitter no pretende situarse como sustitutivo de métodos anteriores, sino que su objetivo es ser capaz de obtener más información de un entorno mucho más dinámico y que no requiere intervención directa por parte de las personas que realizan el estudio.

En el punto presente y que concluimos con este resumen, nuestra misión ha sido justificar y realzar el uso del RAT en un caso particularmente complejo como el nuestro. El RAT permite analizar con precisión grandes cantidades de información fragmentada, dinámica y dispersa (como es el caso de las conversaciones de Twitter), segmentándola, resumiéndola y ordenándola según los criterios que le queramos asignar. Hemos redundado en el nexo entre el RAT y los criterios subjetivos (propios de los métodos cualitativos) y sobre la necesidad de evaluar la calidad de sus resultados usando diversos métodos. Más allá de eso, hemos enfatizado la utilidad del RAT para detectar sentimientos concretos y opiniones dentro de las conversaciones, información que podrá permitir a las empresas, instituciones o entidades el

conocimiento de qué sienten y piensan sus seguidores, clientes o usuarios acerca de los servicios que ofrecen o de las personas que están al frente del proyecto. Si inicialmente hemos situado los *insight* como la expresión (¡eureka!) de una problemática, curiosidad o necesidad latente del sujeto que no es capaz de expresar de forma consciente y explícita por sí mismo, los publicistas tendrán mayores probabilidades de hallarlo si conocen en profundidad todo lo que el público piensa acerca de lo que ellos aspiran a mejorar o reformular de manera novedosa.

En consecuencia, el punto 5 que cierra este capítulo está exclusivamente dedicado al *Sentiment Analysis* y al *Opinion Mining*. Según nuestro parecer, como hemos avanzado anteriormente en este mismo punto, ambos métodos alcanzados gracias al uso del RAT culminan las herramientas que nos permitirán llevar a cabo con éxito nuestra metodología de análisis específica para el hallazgo de *insights* en Twitter.

5. SENTIMENT ANALYSIS

5. *Sentiment Analysis*

En este último punto del segundo capítulo perfilaremos las características y el desarrollo de las mencionadas técnicas de análisis de sentimientos y opiniones. El análisis de sentimientos lo realizaremos con la técnica *Sentiment Analysis* (SA) y haremos lo mismo con las opiniones aplicando la técnica *Opinión Mining* (OM). Avanzaremos que, por cuestiones metodológicas, ambos sentimientos se realizan con el mismo procedimiento y sentimientos y opiniones quedan englobados en el mismo concepto. Que la influencia de las emociones afecta a todas las decisiones que tomamos en nuestro día a día es algo que el psiquiatra y ensayista Daniel Goleman ya testificó a mediados de los noventa en su libro *Inteligencia Emocional*^[370]. Dicha afirmación se amparaba en diversas investigaciones anteriores y en las suyas propias. Desde entonces, una cifra extraordinaria de estudios e investigaciones han ido corroborando, asentando y profundizando el conocimiento acerca de la relación entre las emociones y las acciones que realizamos cotidianamente. El éxito del libro de Goleman contribuyó a la expansión del concepto y, en efecto su influencia no ha dejado de crecer desde entonces y se ha asentado en el lenguaje popular. El interés desde el campo de la publicidad por tratar de dilucidar qué emociones o sentimientos (felicidad, envidia, indignación, bienestar, etc.) promovidos desde los anuncios despiertan un mayor consumo de los productos según su tipología (ya sea un perfume, un coche o un postre dulce), el target (joven, urbano, nicho específico, etc.) y otras variables, no ha parado de crecer y sigue siendo un tema de constante análisis. Para lograr ese conocimiento, ¿de qué mejor manera podríamos obtenerlo que analizando lo que expresan los propios usuarios o consumidores con absoluta naturalidad? Las técnicas *Sentiment Analysis* y *Opinion Mining* tratan justamente de eso en el terreno del análisis de los textos escritos en internet y lo hacen, además, con una mejora sustancial respecto a los métodos tradicionales y que es uno de los puntos que ha abierto esta tesis: la presencia de la persona o grupo de personas que realizan el estudio será ahora inapreciable, con lo que no alterará de ningún modo la realización del mismo ni condicionará el comportamiento de los sujetos analizados. El estudio de las conversaciones, actualmente extendido a través de las redes sociales, foros de opinión, comentarios en medios de comunicación, etc., nos brindará enormes cantidades de datos que las nuevas técnicas serán capaces de convertir en material de enorme utilidad.

El *Sentiment Analysis* obtiene y analiza los sentimientos, opiniones y la orientación de la población en relación a determinados acontecimientos y marcas. En cuanto a la identificación de sentimientos, en nuestro caso optaremos, como hemos apuntado en el primer capítulo, el modo en el que expresará los resultados señalará la polaridad de un documento de texto según la opinión de su creador. (Pang, 2008)^[331] (Musto, 2014)^[371] (Liu, 2012)^[335]

A primera vista, puede resultar sorprendente que antes del año 2000 la técnica del *Sentiment Analysis* se mantuviese prácticamente inexplorada, considerando en especial que se trata de una técnica que “simplemente” analiza textos escritos y procede del PLN. La explicación se halla en la constatación de que, antes del año 2000, apenas había registro de opiniones personales en formato digital, de modo que el material básico de análisis era tan escaso que no estaba justificado emprender su estudio. Sin embargo, el cambio se ha producido dramáticamente con el *boom* de la web y la masificación de los medios sociales. En apenas unos años, investigadores de muy diversos campos se han encontrado delante de una cantidad gigantesca de datos a su alcance y con la posibilidad de acceder a ellos sin limitación (Liu, 2015)^[372]. El campo de la publicidad se ha visto beneficiado directamente con esta nueva opción, y el consumidor y su psicología se han acercado de manera natural y explícita a las necesidades y objetivos que el trabajo publicitario aspira a resolver. La vía que permitirá la expresión del contenido publicitario y la saciedad de las expectativas, deseos o necesidades de los consumidores pasa indefectiblemente por la generación de *insights* apropiados. Esperamos que el cierre de este capítulo nos confirme la pertinencia de aplicar las nuevas técnicas SA y OM para alcanzar nuestra misión como publicistas e investigadores.

5.1. Antecedentes

El término *sentiment* en el contexto del análisis de textos fue mencionado por primera vez en el documento *Yahoo! For Amazon: Opinion Extraction from Small Talk on the Web* elaborado por Sanjiv R. Das y Mike Y. Chen. Su objetivo con la detección de sentimientos responde a las necesidades que a día de hoy siguen siendo las más demandadas en el campo de las finanzas, intentar medir de algún modo la confianza de los inversores: “We present a methodology for real time sentiment extraction in the domain of finance. [...] This paper comprises (a) a technology for extracting small investor sentiment from web sources to create an index, and (b) illustrative applications of the methodology”. (Das, 2001:2)^[373]

Richard Tong, durante una conferencia en New Orleans en octubre de 2001 sobre *Operational Text Classification*, propuso una nueva tarea de investigación consistente en la detección, seguimiento y monitorización de las opiniones desarrolladas en entornos digitales para la posterior clasificación de los sentimientos y opiniones. (Tong, 2001)^[374]

Sin embargo, uno de los equipos más prolíficos en este campo es el que forman Bo Pang y Lillian Lee. Ya en el año 2002 centraron sus intereses en el análisis de los sentimientos expresándolo de la siguiente manera: “recent years have been seen rapid growth in on-line discussions groups and review sites, where a crucial characteristic of the posted articles is their sentiment, or overall opinion towards the subject matter - for example, whether a product review is positive or negative.” (Pang, 2002:79)^[375]. Su implicación desde ese momento en el desarrollo del *Sentiment Analysis* ha venido marcada por la toma de consciencia de la relevancia e influencia de las valoraciones y opiniones en el comportamiento futuro de aquellos que deberán decidir sobre lo mismo en un futuro (como las valoraciones de los clientes que ya han comprado ese producto y califican el producto según su satisfacción).

A finales de 2002, Peter Turney presentaba un nuevo algoritmo para clasificar los comentarios de los usuarios de internet en función de si recomendaban positivamente el contenido (o producto) o desaconsejaban su consumo. El autor lo describió como *thumbs up*

(pulgares arriba, equivalente a recomendación positiva, *thumbs down*, pulgares abajo, equivalente a consumo desaconsejado). Este método se extendería de forma masiva en los próximos años y, de hecho, es uno de los emblemas más conocidos de Facebook (el *thumbs up* o “me gusta”) en la actualidad. Tal y como lo indicó el mismo Turney: “This paper presents a simple unsupervised learning algorithm for classifying reviews as recommended (thumbs up) or not recommended (thumbs down). The classification of a review is predicted by the average semantic orientation of the phrases in the review that contain adjectives or adverbs”. (Turney, 2002:417)^[376]

En 2003, Tetsuya Nasukawa y Jeonghee Yi profundizaban en el análisis de los sentimientos asociados a la polaridad siguiendo el esquema “positivo o negativo”. Introdujeron un criterio interesante que permitía identificar los sentimientos en relación a fragmentos concretos y no únicamente en la totalidad del texto. De este modo, lograron incrementar los matices y oscilaciones dentro de la opinión general del usuario: “The essential issues in sentiment analysis are to identify how sentiments are expressed in texts and whether the expressions indicate positive (favorable) or negative (unfavorable) opinions toward the subject. In order to improve the accuracy of the sentiment analysis, it is important to properly identify the semantic relationships between the sentiment expressions and the subject. By applying semantic analysis with a syntactic parser and sentiment lexicon, our prototype system achieved high precision (75-95%, depending on the data) in finding sentiments within Web pages and news articles.” (Nasukawa, 2003:70)^[377]

Tras el primer trabajo en el campo del *Sentiment Analysis* por parte de Tetsuya, él mismo presentó y junto a otros tres investigadores un *Sentiment Analyzer*. El sistema permitía la extracción de sentimientos acerca de un tema concreto, utilizando técnicas del PLN en las que consideraban las opiniones equivalentes a los sentimientos.

En 2003 destacó también el trabajo conjunto de Yu y Hatzivassiloglou. Su trabajo presentaba un modelo para la clasificación de la subjetividad a partir de la similitud de las frases. Sus premisas partían de la consideración de que las frases de carácter subjetivo (expresan sentimientos y opiniones) son más similares entre sí que las frases objetivas. (Yu, 2003)^[378]

En cuanto a la evolución de la técnica *Opinion Mining*, encontramos una mención por primera vez en el informe de Kushan Dave, Steve Lawrence y David Pennock hecho ex profeso para la conferencia *Proceedings of the 12th international conference on World Wide Web* que tuvo lugar en Budapest el 2003. En su descripción iniciática del concepto aludían a su popularidad entre los usuarios intensivos de búsquedas en internet o la mencionaban en referencia a la recuperación de la información (Dave, 2003)^[379]. Según sus autores, la OM ideal sería aquella capaz de “procesar un conjunto de resultados de búsqueda para un elemento concreto, generando una lista de productos (atributos de calidad, características, etc.) junto a las opiniones correspondientes sobre cada uno de ellos (malas, buenas, mezcladas)” (Dave, 2003:519)^[379]. Una gran parte de la investigación realizada con posterioridad al planteamiento inicial de los tres investigadores, se centrará más en los aspectos relacionados con la extracción y el análisis de las opiniones acerca de los diferentes aspectos de los objetos analizados. Más recientemente se ha ampliado la interpretación del término y se han incorporado otros aspectos del análisis de textos. (Liu, 2007)^[380]

El campo del OM recibiría una novedosa aportación de la mano de los investigadores Pang y Lee. En 2005 presentaron la clasificación de comentarios por número de estrellas y que en muy poco tiempo se extendería y se convertiría en una herramienta de uso común en las webs que ofrecen servicios, productos o contenidos “evaluables”, y muy en particular en las webs dedicadas al comercio electrónico.

En su propuesta, titulada *Seeing Stars: exploiting class relationships for sentiment categorization with respect to rating scales*, Pang y Lee apostaban por un sistema que profundizaba en la granularidad del análisis. En lugar de limitarse a la información que comunica el *thumbs up* o *thumbs down*, proponían una calificación numérica creada por el autor que pudiese ser expresada en un número determinado de estrellas (tres estrellas, cinco estrellas, etc.). (Pang, 2005)^[381]

5.2. Definición

Tal y como argumentamos en el capítulo de esta tesis correspondiente a la terminología utilizada, seguiremos principalmente el concepto y visión de *Sentiment Analysis* y *Opinion Mining* que sostienen los investigadores Bing Liu, Bo Pang y Lei Zhang y en las siguientes definiciones utilizaremos literalmente algunos de sus textos:

- Los términos *Sentiment Analysis* y *Opinion Mining* se presentan en un gran número de ocasiones como sinónimos. Sin embargo, mostraremos en un punto de las contribuciones a las diversas definiciones una propuesta para señalar la variación conceptual entre ambos términos.
- Para una persona, la labor de búsqueda de los sitios en Internet con los contenidos deseados, la obtención y resumen de dicha información es una labor inalcanzable. Además, los análisis de textos realizados por personas son más fácilmente susceptibles de estar sujetos a sesgos (Kahneman, 2011)^[382]. Es imprescindible entender esta situación para hacernos cargo de la necesidad del uso de sistemas automáticos en cuanto a los SA.

La definición inicial más completa que ofrecemos de *Sentiment Analysis* (SA) y *Opinion Mining* (OM) reúne las aportaciones por separado de los tres investigadores que hemos situado como referentes en este campo de estudio.

Bing Liu ofrece una definición de carácter más descriptivo acompañándola de otros términos sinónimos que también definirían el concepto. En el segundo párrafo hace hincapié en su importancia y utilidad con la mención de algunos campos concretos en los que puede aplicarse el *Sentiment Analysis* y obtener resultados satisfactorios:

Sentiment Analysis, also called “Opinion Mining”, is the field of study that analyzes people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space. There are also many names and slightly different tasks, e.g., “sentiment analysis”, “opinion mining”, “opinion extraction”, “sentiment mining”, “subjectivity analysis”,

“affect analysis”, “emotion analysis”, “review mining”, etc.. However, they are now all under the umbrella of sentiment analysis or opinion mining. (Liu, 2012:7)^[335]

Sentiment Analysis or Opinion Mining is the computational study of people’s opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes. The task is technically challenging and practically very useful. For example, businesses always want to find public or consumer opinions about their products and services. Potential customers also want to know the opinions of existing users before they use a service or purchase a product. (Liu, 2012:415-416)^[335]

Si la definición de Liu ponía el ámbito empresarial como ejemplo central para aplicar la técnica, la de otro investigador, Zhou, amplia los ejemplos a un campo mucho más abierto y variable como es el de los sucesos o eventos (políticos, económicos, fenómenos, etc.).

Opinion mining and sentiment analysis can be described as the process of automatically extracting and analyzing the opinions, sentiments, thoughts and feelings of opinion writers on a specific target. This target could be a product, some issue like politics, economics, events, phenomena, services, etc. (Zhou, 2006:214)^[359]

Como señalábamos, para autores como Bo Pang, *Sentiment Analysis y Opinion Mining* operan en el mismo ámbito y en sus textos se refieren a ambos términos como sinónimos, pero citaremos la definición que ofrece Bing Liu de *Opinion Mining* para resaltar las diferencias que él sí establece entre ambos conceptos:

Given a set of evaluative text documents D that contain opinions (or sentiments) about an “object” (person, organization, product etc.), opinion mining aims to extract attributes and components of the object that have been commented on in each document d in the set D and to determine whether the comments are positive, negative or neutral. (Liu, 2009:1986)^[349]

Como apuntamos al hablar de SA en el punto anterior (2.4) y queremos remarcar ahora que hemos profundizado en su definición, es importante tener en cuenta la dificultad que presenta su aplicación por el hecho de que su materia de estudio es el lenguaje natural. El lenguaje natural, entre otras caracterizaciones, es totalmente dinámico (evoluciona día a día con su uso en la sociedad) y presenta una alta variabilidad semántica. La polisemia, el doble sentido, los neologismos, el lenguaje coloquial, todo ello son factores que dificultan el SA. Gracias a los

grandes avances y a la creciente sofisticación de los algoritmos, las mejoras son constantes y, si bien la perfección no es alcanzable, los resultados obtenidos desde los inicios hasta ahora son cada vez mejores y probadamente satisfactorios, fiables y relevantes.

5.2.1. Emociones, sentimientos y opiniones

Hemos definido en SA que, por una cuestión de claridad y para evitar la dispersión terminológica, y siempre con el fin de lograr un análisis igualmente satisfactorio, las emociones, los sentimientos y las opiniones se engloban dentro del concepto *sentiment*. La clave del sentido de esta agrupación es que, si realizamos la simplificación de emociones, sentimientos y opiniones a las categorías de positivo/negativo, podemos entender que expresar una opinión, un sentimiento o una emoción de un tipo u otro apunta a una determinada visión coherente. Es decir, si una persona siente y expresa en Twitter una emoción de “rabia”, difícilmente añadirá que opina que aquello que le provoca “rabia” le parece una “buena” cosa. El trabajo realizado con esta conjunción de términos ha probado reflejar con alta corrección la pertinencia de su aplicación en SA. Ello no implica que, tanto académicamente como para nuestro trabajo, nos parezca menos relevante dedicar un apartado a definir por separado y de manera más completa qué entendemos por emoción, sentimiento y opinión separadamente.

Como punto de partida, la definición que ofrece el diccionario *Merriam-Webster* de sentimiento nos refiere a “una actitud, pensamiento o juicio”, mientras que define opinión como una “visión, juicio o valoración formada sobre un asunto, en la mente de un particular”. En este caso, la diferencia que nos ofrece es bastante sutil. Con ambas definiciones podríamos apuntar que una opinión indica más el punto de vista acerca de algo y que se presta más a ser fruto de una valoración que incluya argumentos. En cuanto a las emociones, podemos señalar que surgen con mayor fuerza y tienden a ser impulsivas. Por ejemplo, la frase: “Estoy preocupado por la crisis económica” expresa un sentimiento, mientras que la frase: “Creo que el ministro no está haciendo bien las cosas”, expresa una opinión. La preocupación puede estar o no justificada, pero en todo caso es algo más profundo que la

expresión de una “creencia”. Para entrar más en detalle y con la mayor precisión posible, lo presentamos seguidamente.⁶⁶

5.2.1.1. Emoción

Las emociones han sido estudiadas con detenimiento desde los filósofos clásicos y su intento por teorizar de un modo organizado aquello que les transmitía su propia experiencia o las reacciones que encontraban en los demás. Más recientemente, la ciencia moderna y la psicología adoptaron nuevas e importantes consideraciones. En la actualidad, la disciplina de la neurociencia es la que está alcanzando mayores niveles de conocimiento al estudiar qué procesos neurológicos generan las emociones y todo aquello que las provoca (De Sousa, 2013)^[383]. Hasta el tardío desarrollo de la ciencia y sus recientes logros en el estudio del funcionamiento del cerebro, el estudio de las emociones ha presentado numerosas divergencias y discusiones, precisamente debido a su intangibilidad, diversidad y a la profunda complejidad del comportamiento del ser humano. A modo de ejemplo, esta definición corresponde a finales del siglo XIX: “The bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur is the emotion.” (James, 1884:189)^[384]. A mediados del siglo XX, el investigador Ryle constataba la complejidad del término: “The word “emotion” is used to designate at least three or four different kinds of things.” (Ryle, 1949:81)^[385]. En el último cambio de siglo, Russell y Barret-Feldman mantenían una posición similar: “Emotion is too broad a class of events to be a single scientific category” (Russell, 1999:805)^[386].

Antes de llegar a las definiciones contemporáneas que consideramos más representativas, mencionaremos las dos escuelas que mantuvieron en la modernidad dos visiones divergentes. Por un lado, la tradición evolucionista, iniciada por Darwin, propone que las emociones están directamente relacionadas con los sistemas que se desarrollaron al tratar situaciones biológicamente significativas y, en consecuencia, tienen un carácter universal. En

⁶⁶ Es necesario observar que *sentiment* y *feeling* se traducen a veces como sinónimos en castellano. Sin embargo, en inglés tienen significados ligeramente diferentes. La totalidad de la bibliografía consultada para esta tesis (a excepción del uso del término *feeling* en las investigaciones realizadas en el campo de la Psicología y la Filosofía) se refiere únicamente a *sentiment* y no a *feeling*. En consecuencia, hemos optado por traducir *sentiment* por “sentimiento”.

contraposición, el constructivismo social se refiere a las emociones como consecuencia de las reglas sociales y culturales. (Averill, 1980)^[387] (Harre, 1989)^[388]

Aportamos a continuación tres importantes definiciones ofrecidas recientemente y que nos acercan a una comprensión vigente de las emociones y de su naturaleza.

Entendemos por emoción aquello que normalmente es producido por una persona que evalúa un evento, consciente o inconscientemente, en tanto que resulta relevante para un objetivo o una meta; una emoción da prioridad a una línea de actuación, interrumpiendo otros procesos mentales, a veces seguida de cambios corporales, expresiones o acciones. (Izard, 1971)^[389]

Los humanos tenemos emociones y después sentimientos y estados de ánimo porque evolutivamente la emoción precede a los sentimientos y a los estados de ánimo. (Damasio, 2010)^[390]

Una emoción es una reacción afectiva que surge súbitamente ante un estímulo, dura un corto espacio de tiempo y comprende una serie de repercusiones fisiológicas, cognitivas y comportamentales. Como características principales se pueden destacar su corta duración, alta intensidad, innatas y universales, provocadas por un estímulo concreto y producen reacciones fisiológicas. (Delgado Linares, 2014:28)^[391]

Por su lado, Scherer consideró la emoción desde un punto de vista más amplio, vinculando el organismo que la tiene y el medio o “entorno” en el que sucede, en constante cambio y en el que el ser humano debe desenvolverse en función de sus incesantes modificaciones.

El proceso y la utilidad de las emociones según Scherer es la siguiente: (Scherer, 1984)^[392]

- Analizar la importancia del estímulo según las necesidades del organismo, sus preferencias, intenciones.
- Preparar físicamente y psicológicamente al organismo para llevar a cabo una acción adecuada.
- Permitir a otros organismos conocer cuales son sus intenciones.

En cuanto a su categorización o tipología de emociones, reunimos diversas propuestas de los investigadores Ekman, Plutchik y Parrot.

Según Ekman, las emociones se pueden agrupar en un conjunto de seis básicas y que puede asociarse a las expresiones faciales: Enfado (*anger*), asco o aversión (*disgust*), miedo (*fear*), alegría (*joy*), tristeza (*sadness*) y sorpresa (*surprise*). (Ekman, 1999:550-5553)^[393]

Otro modelo más detallado de clasificación lo encontramos en la “Rueda de la emociones” de Plutchik, quien propone las siguientes emociones: *joy, trust, fear, surprise, sadness, disgust, anger* y *anticipation*.

Como se verá en el gráfico, la dimensión vertical del cono representa la intensidad de cada tipo y cada uno de los ocho sectores del círculo describe un tipo de emoción primaria. La rueda de las emociones se complementa a través de una representación de colores suplementarios (adyacentes) y complementarios (opuestos). Las emociones del lado opuesto del círculo representan la emoción complementaria. Asimismo, queda reflejado en el gráfico que las emociones también se pueden mezclar hasta obtener diadas de emociones, como por ejemplo: disfrutar (*joy*) + confianza (*trust*) = amor (*love*). (Plutchik, 1991:111)^[394]

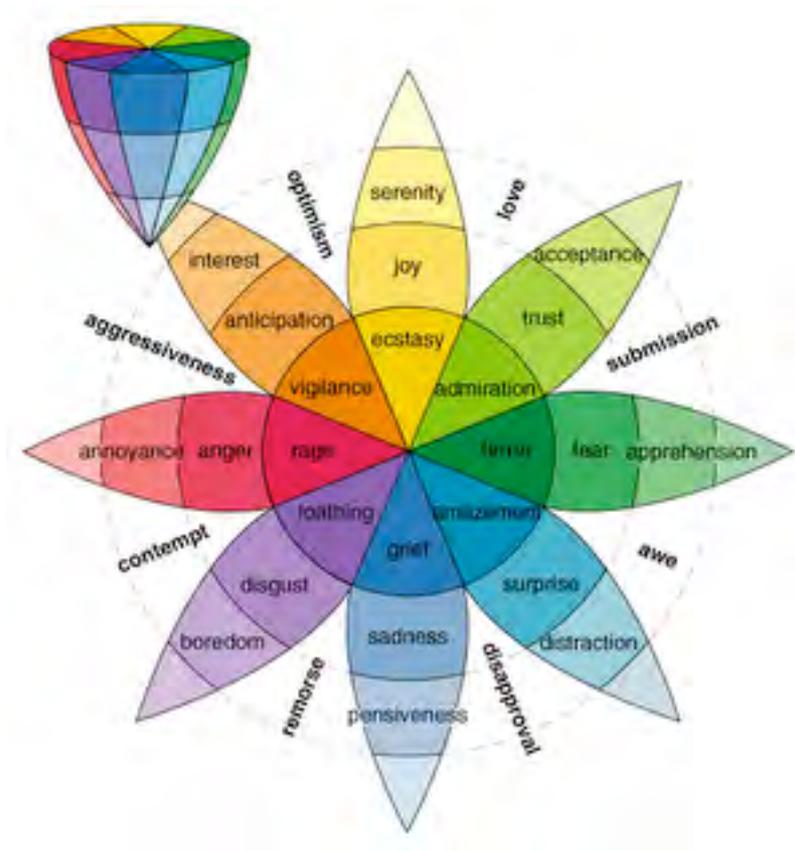


Figura 45: La Rueda de las emociones de Plutchik. (Plutchik, 2001:349)^[395]

Parrott, más recientemente, desarrolló una tabla que presenta en tres columnas, dividiendo las emociones según primarias, secundarias y terciarias. El total de emociones que propone es de 150.

Primary emotion	Secondary emotion	Tertiary emotions
Love	Affection	Adoration, affection, love, fondness, liking, attraction, caring, tenderness, compassion, sentimentality
	Lust	Arousal, desire, lust, passion, infatuation
	Longing	Longing
Joy	Cheerfulness	Amusement, bliss, cheerfulness, gaiety, glee, jolliness, joviality, joy, delight, enjoyment, gladness, happiness, jubilation, elation, satisfaction, ecstasy, euphoria
	Zest	Enthusiasm, zeal, zest, excitement, thrill, exhilaration
	Contentment	Contentment, pleasure
	Pride	Pride, triumph
	Optimism	Eagerness, hope, optimism
	Enthrallment	Enthrallment, rapture
	Relief	Relief
Surprise	Surprise	Amazement, surprise, astonishment
Anger	Irritation	Aggravation, irritation, agitation, annoyance, grouchiness, grumpiness
	Exasperation	Exasperation, frustration
	Rage	Anger, rage, outrage, fury, wrath, hostility, ferocity, bitterness, hate, loathing, scorn, spite, vengefulness, dislike, resentment
	Disgust	Disgust, revulsion, contempt
	Envy	Envy, jealousy
	Torment	Torment
Sadness	Suffering	Agony, suffering, hurt, anguish
	Sadness	Depression, despair, hopelessness, gloom, glumness, sadness, unhappiness, grief, sorrow, woe, misery, melancholy
	Disappointment	Dismay, disappointment, displeasure
	Shame	Guilt, shame, regret, remorse
	Neglect	Alienation, isolation, neglect, loneliness, rejection, homesickness, defeat, dejection, insecurity, embarrassment, humiliation, insult
Fear	Horror	Alarm, shock, fear, fright, horror, terror, panic, hysteria, mortification
	Nervousness	Anxiety, nervousness, tenseness, uneasiness, apprehension, worry, distress, dread

Tabla 4: Las emociones primarias, secundarias y terciarias según Parrot. (Parrott, 2001)^[396]

Retomando la consideración de *sentiment* bajo el que el SA engloba las emociones, la aclaración y visión descrita por Cowie nos resulta especialmente significativa y responde al cómo puede estudiarse una emoción, no desde la expresión facial que genera u otras observaciones, sino desde el texto escrito:

La palabra emoción ha sido analizada desde diversas disciplinas, siendo la psicología una de las principales, pero dado que este estudio se centra en el análisis de los sentimientos expresados por los usuarios en forma de textos, consideramos la descripción de la emoción desde la perspectiva del discurso, lo que algunos investigadores denominan como: *speech and emotion*. (Cowie, 2003:5)^[397]

5.2.1.2. Sentimientos

Los sentimientos engloban una gran variedad de aspectos que van desde las sensaciones físicas más básicas como el tacto o el dolor, hasta complejas emociones como el miedo, el odio, la tristeza, la felicidad, entre otras. En Psicología, este concepto se usa frecuentemente para comprender la experiencia consciente y subjetiva de las emociones, tal y como indica el investigador Van den Bos en el *APA Dictionary of Psychology*, quien define el término sentimiento como “the conscious subjective experience of emotion” (Vandenbos, 2007)^[398]. El investigador portugués Antonio Damasio considera al sentimiento como la percepción de lo que sucede en nuestro cuerpo y mente cuando se manifiestan emociones (Damasio, 2010)^[390].

Los sentimientos son procesos afectivos relativamente estables, adquiridos en el proceso de la socialización, experimentados por seres humanos. Como ejemplos de sentimientos tenemos el amor hacia nuestra pareja o el odio a nuestros enemigos. Características principales: son subjetivos, surgen de forma lenta y progresiva y promueven conductas. (Francisco, 2008:11)^[399]

5.2.1.3. Opiniones

La orientación de una opinión sobre una función u objeto indica si la opinión es positiva, negativa o neutra. La orientación de la opinión también se conoce como: “*sentiment orientation*”, “*polarity of opinion*” o “*semantic orientation*”. Una opinión sobre una función u objeto es una visión positiva o negativa, actitud, emoción o evaluación de un titular de opinión. El titular de la opinión es la persona u organización que expresa la opinión. Una opinión explícita sobre una función u objeto, es una opinión expresada explícitamente en una frase subjetiva, mientras que una opinión implícita sobre una función u objeto, es una opinión expresada en una frase objetiva. Una frase con opinión es una frase que expresa de forma explícita o implícita opiniones positivas o negativas y puede ser una frase subjetiva u objetiva. (Liu, 2012:4-5)^[335]

Todas las tareas definidas dentro del ámbito de la minería de opiniones “*Opinion Mining*” para la clasificación de textos, se desarrollan según la orientación de la opinión, que por lo general, se clasifican en tres clases: positiva, negativa o neutra. (Pang, 2008:16)^[331]

Las opiniones tienen tres componentes básicos: El titular de la opinión: que corresponde a la fuente de la opinión y que puede ser una opinión particular o de una organización. El objeto de análisis: que corresponde a la expresión de una opinión sobre una función o característica del objeto evaluado y la opinión en sí misma: que corresponde a la orientación semántica de una opinión, sobre una función de un objeto con una valoración positiva, negativa o neutra. (Liu, 2012:15)^[335]

El análisis de los sentimientos se puede dividir en distintas sub-tareas: subjetividad - Objetividad: que determina si se está expresando un hecho probado o simplemente una opinión. Polaridad positiva o negativa: que determina si el texto subjetivo que se está analizando expresa una opinión positiva o negativa. La fuerza de la opinión: que determina si se trata de una opinión débilmente positiva/negativa o fuertemente positiva/negativa o neutra. Emoción: que sirve para determinar la emoción concreta que se está tratando de transmitir. Determina si se trata de un texto alegre, triste, etc. (Francisco, 2008:8)^[399]

5.3. Tipología de métodos del *Sentiment Analysis*

Los métodos existentes de SA son muy diversos y responden a la gran variedad de necesidades que pueden presentar los distintos estudios en los que puede aplicarse. Sin embargo, existe una división principal entre dos tipos de método cuyo modo de clasificar el texto y sus partes determinará enormemente los resultados obtenidos. Es preciso conocer las ventajas e inconvenientes de ambos para acertar correctamente en su aplicación.

La principal división de los métodos de SA los separa entre los métodos *Machine Learning* (ML), los basados en el léxico (Gonçalves, 2013)^[400] y, en un tercer lugar, encontramos aquellos que son un híbrido entre ambos (Maynard, 2011)^[401]. Los métodos ML suelen clasificar los datos con supervisión y de forma binaria (positivos o negativos). Como veremos, su principal desventaja es que, para etiquetar la información y clasificarla (Gonçalves, 2013)^[400] (Pang, 2002)^[375] antes de que el sistema pueda aprenderlo de forma automática, se requiere un entrenamiento previo de los supervisores que exige tiempo y presenta un coste elevado.

Como ramificación de la Inteligencia Artificial, el ML trabaja con el diseño y desarrollo de algoritmos que permiten a los ordenadores adoptar patrones de comportamiento basados en datos empíricos. Una de las aspiraciones del ML es que los ordenadores aprendan a reconocer patrones complejos automáticamente y tomen decisiones inteligentes con los datos que relacionan y siempre partiendo de una serie de datos etiquetados con anterioridad. (Hemalatha, 2013:106)^[402] (Pak, 2010)^[403] (Hu, 2013)^[404]

En cuanto a la conveniencia de su aplicación, los métodos *learning-based* como ML tienen la ventaja de poder aplicarse en modelos específicos y contextos concretos pero, como apuntábamos, su principal inconveniente es que el “entrenamiento” que requiere el etiquetaje previo de los datos (y la dificultad de etiquetar nuevos en un futuro) suponen un tiempo y un coste que puede resultar, según el tamaño del volumen de información que se aspira a clasificar, demasiado costoso o directamente inasumible. (Gonçalves, 2013)^[400]

Por contra, los métodos basados en el léxico utilizan un set predeterminado (externo) de palabras y patrones (normalmente conocido como *sentiment dictionary* o *lexicon*) con cada

componente asociado a un determinado sentimiento dentro de una categoría (positivo, negativo, neutro) o una puntuación numérica que permite medir su intensidad o fuerza. Al usar este diccionario ya diseñado, no requiere un etiquetaje previo ni, en consecuencia, tampoco necesita un aprendizaje o “entrenamiento”. El reto recae en aplicar un *lexicon* de calidad y adecuado al tipo de texto que se quiere analizar (Musto, 2014)^[371]. En este sentido, suele utilizarse en los análisis del lenguaje mediático y es menos válido o útil en el caso del lenguaje coloquial (precisaría un etiquetaje previo y concreto, es decir, sería más recomendable aplicar un método ML). (Shalunts, 2014)^[405]

5.3.1. Tipología de métodos basados en el *Machine Learning* (ML)

Recogemos y definimos los rasgos esenciales de los métodos ML más utilizados en los SA.

5.3.1.1. *Naïve Bayes*

Se conoce por ser el clasificador más simple y de uso más común. El modelo trabaja con el estudio de la posición de las palabras en el texto empezando por la aplicación del modelo *Bag of Words* (capaz de hacer el estudio de forma contraria, sin tener en cuenta el orden). Gracias al Teorema de Bayes puede predecir la probabilidad de que un determinado conjunto de palabras pertenezcan a las etiquetas concretas definidas con anterioridad. (Medhat, 2014:1099)^[406]

5.3.1.2. *Maximum Entropy*

El clasificador Maxent (*Maxent Classifier* o *conditional exponential classifier*) convierte las etiquetas en vectores mediante la codificación de las mismas. Los vectores codificados pueden ser utilizados para calcular los “pesos” de las etiquetas y con ello y el texto se puede encontrar las relaciones más probables. Los parámetros se obtienen señalando una determinada cantidad de X (peso) y su combinación con la correspondiente cantidad de X (codificada). (Medhat, 2014:1100)^[406]

5.3.1.3. Support Vector Machines (SVM)

La función principal de los SVM es determinar separadores lineales en los cuadros de búsqueda según diversas clases. Los textos son idóneos para la clasificación SVM: pese a su disparidad, sus características son limitadas, suelen estar correlacionados y generalmente se presentan organizados en categorías separables linealmente (Joachims, 1996)^[407]. Los SVM se aplican en casos muy diversos, por ejemplo en la clasificación de reseñas según su calidad (Chen, 2011)^[408] o en la separación por polaridades de los sentimientos. (Li, 2013)^[409]

5.3.2. Tipología de métodos basados en el léxico (*lexicon-based*)

Recogemos y definimos los rasgos esenciales de los métodos basados en el léxico más utilizados en los SA.

5.3.2.1. Cálculo de Orientación Semántica (*Semantic Orientation Calculator*) (SOCal)

La Orientación semántica permite medir la subjetividad y la opinión vertidas en un texto. Suele hallar un factor valorativo (positivo o negativo) y una potencia o fuerza (la “intensidad” o “grado” de positividad o negatividad de la palabra, frase o texto) dirigidos hacia un tema, una persona o una idea (Osgood, 1957)^[410]. Su aplicación en el análisis de la opinión pública o en las interpretaciones de las reseñas de productos *on-line* resulta muy útil en el campo del marketing y en los estudios de popularidad. (Taboada, 2011:268)^[336]

El cálculo de Orientación semántica utiliza diccionarios con palabras que incorporan su orientación semántica (valoración positiva/negativa y su potencia o fuerza), la intensidad o grado y también su negación. Su aplicación habitual corresponde a la tarea de clasificación por polaridad (el proceso de asignar una etiqueta positiva o negativa a un texto que resume representa la opinión principal respecto al tema del que habla). (Musto, 2014)^[371]

5.3.2.2. WordNet

Amplia base de datos con léxico en inglés (nombres, verbos, adjetivos y adverbios) creada originalmente por George Miller, profesor de psicología de la Universidad de Princeton (Miller, 1995)^[411]. Su interés reside en que las palabras están enlazadas según una serie de criterios léxicos y de relaciones conceptuales que forman una “red” en su conjunto. WordNet hace especial énfasis en los sinónimos (a los que llama *synsets*). Los *synsets* están enlazados entre sí por significados conceptuales y por sus relaciones léxicas. El resultado de esta red de *synsets* puede ser consultable usando un navegador.

Aunque el método no ha sido significativamente desarrollado en SA, sí se ha utilizado para procesos de análisis de texto que guardan cierta relación con nuestra materia. En la siguiente definición hablaremos de un recurso particular especialmente destinado al SA y basado en WordNet. Además, WordNet es accesible y descargable de manera gratuita. Su estructura hace el programa especialmente útil para la lingüística computacional y el PLN. (Fellbaum, 2005)^[412]

Semantic Relation	Syntactic Category	Examples
Synonymy (similar)	N, V, Aj, Av	pipe, tube rise, ascend sad, unhappy rapidly, speedily
Antonymy (opposite)	Aj, Av, (N, V)	wet, dry powerful, powerless friendly, unfriendly rapidly, slowly
Hyponymy (subordinate)	N	sugar maple, maple maple, tree tree, plant
Meronymy (part)	N	brim, hat gin, marini ship, fleet
Troponymy (manner)	V	march, walk whisper, speak
Entailment	V	drive, ride divorce, marry

Note: N - Nouns Aj - Adjectives V - Verbs Av - Adverbs

Figura 46: Relaciones semánticas en WordNet. (Miller, 1995:40)^[411]

Existen dos extensiones de *WordNet* recomendables para el SA:

- *SentiWordNet*. (Baccianella, 2010)^[413] Es un recurso léxico que sirve como complemento a los sistemas de SA. Para cada *synset* de *WordNet*, ofrece una anotación basada en tres puntuaciones numéricas de los sentimientos (positividad, negatividad, neutralidad) (Miller, 1995)^[411]. Nos interesa remarcar que *SentiWordNet*, al estar basado en una red de *synsets*, es capaz de dar una puntuación diferente a la misma palabra en función del sentido con el que se haya vinculado en cada ocasión.
- *WordNet-Affect*. (Strapparava, 2004)^[414] Es una herramienta lingüística que permite una representación de los afectos. Como extensión de *WordNet* permite etiquetar *synsets* con conceptos afectivos denominados A (por ejemplo, el término “euforia” se asocia al concepto “emoción positiva”, el nombre “enfermedad” se asocia al concepto “estado físico”, etc.). (Musto, 2014)^[371]

5.3.2.3. *MPQA Subjectivity Lexicon*

Proporciona una base léxica de 8.222 términos (etiquetados como “expresiones subjetivas”) provenientes de distintas fuentes. Las palabras que contiene incorporan la polaridad (positiva, negativa, neutral) y la “intensidad” (fuerte, débil). (Musto, 2014)^[371] (Wiebe, 2005)^[415]

5.3.2.4. *SenticNet*

SenticNet (Cambria, 2014)^[416] es un recurso léxico para los SA. Está basado en la *Sentic Computing* (Cambria, 2012)^[417] un nuevo sistema computacional basado en un paradigma multidisciplinario del SA. A diferencia de los recursos presentados previamente, *SenticNet* es capaz de asociar la polaridad y la información afectiva con conceptos como alcanzar un objetivo o celebrar una ocasión especial, entre otros. En la actualidad, *SenticNet* ofrece sentimientos puntuados (en un rango entre -1 y 1) para 14.000 conceptos de uso habitual. Cada sentimiento asociado a un término se define según la intensidad de dieciséis emociones básicas, definidas a partir del modelo llamado *Hourglass of Emotions*.



Figura 47: Visión en 3D del modelo Hourglass of Emotions y la red Hourglass of Emotions. Recibe el nombre de Hourglass (reloj de arena) debido a la forma resultante de la red al haber relacionado los estados afectivos con su fuerza (desde muy positiva a muy negativa). (Cambria, 2012:152)^[417]

5.3.2.5. Linguistic Inquiry and Word Count (LIWC)

El LIWC⁶⁷ es un programa de análisis de textos diseñado por James W. Pennebaker, Roger J. Booth y Martha E. Francis. Calcula el uso que realizan las personas de diferentes categorías de palabras desde diversas fuentes de texto, pudiendo incluir correos electrónicos, discursos, artículos o la transcripción de conversaciones cotidianas. De forma inmediata, analiza los contenidos de cualquier texto determinando el uso de emociones positivas o negativas, auto referencias, palabras causales y hasta un total de 82 dimensiones lingüísticas. El programa LIWC puede analizar cientos de archivos de texto ASCII estándar o documentos de Microsoft Word en cuestión de segundos. La versión LIWC2007 permite crear diccionarios

⁶⁷ Fuente: <http://www.liwc.net/> (201403)

propios afinar de este modo el análisis de términos del lenguaje relevantes para los objetivos precisos de cada estudio. (Pennebaker, 2001)^[27]

El propósito de su creación es proporcionar un método eficiente y eficaz para el estudio de los componentes de los distintos procesos estructurales presentes en las manifestaciones verbales y escritas de los individuos, sean estas emocionales o cognitivas. La segunda versión (LIWC2007) es capaz de analizar el texto escrito, palabra por palabra, calcular el porcentaje de las palabras en el texto que coinciden con cada uno alguna de las 82 dimensiones del lenguaje existentes y generar un informe en forma de archivo de texto delimitado por tabuladores y legible directamente en los programas de la aplicación, SPSS para Windows, Excel, entre otros (Pennebaker, 1996)^[25]. También permite realizar búsquedas dentro del propio programa y analizar los textos a partir de las frases. En la versión actual LIWC2007 pueden realizarse búsquedas y analizar textos a partir de frases. (Ver Anexo: Diccionario LIWC).

5.3.3. Los diferentes niveles de clasificación en el Sentiment Analysis

Encontraremos hasta tres niveles de clasificación del contenido de los textos en el uso del *Sentiment Analysis* (Medhat, 2014:1094)^[406]. Como quedará patente, no son incompatibles entre sí y dependen de la cantidad de texto de la que quiera extraerse información de una sola vez. A mayor cantidad de texto (documento completo) la precisión se pierde considerablemente pero se alcanza una conclusión general que puede resultar útil en ciertos casos. En el nivel más detallado (nivel de aspecto) logramos detectar y precisar significativamente el peso de cada sentimiento.

5.3.3.1. Nivel del texto (documento) completo

La tarea a realizar en nivel más amplio es la calificación de la opinión general del texto completo en función de si expresa un sentimiento positivo o negativo (Pang, 2002)^[375] (Turney, 2002)^[376]. Por ejemplo, dada una reseña de un producto, el sistema determinará si es una opinión mayormente positiva o negativa respecto al reproductor. en este nivel de análisis se asume que cada documento expresa opiniones respecto a una entidad determinada (por

ejemplo, un único producto). En consecuencia, no será aplicable a documentos que evalúan o comparan diversos elementos. (Liu, 2012)^[335]

5.3.3.2. Nivel de las frases

En este segundo nivel de detalle, el objetivo es clasificar el sentimiento expresado en cada frase. El primer paso consiste en determinar si la oración es subjetiva u objetiva. Si es subjetiva, en este nivel se podrá determinar si la frase expresa opiniones positivas o negativas. Sin embargo, una clasificación a nivel del documento completo o a nivel de las frases que contiene no proporciona el detalle de las opiniones necesario para un gran número de casos y análisis. Para ello será preciso entrar en el nivel de aspecto. (Medhat, 2014)^[406]

5.3.3.3. Nivel del aspecto de las palabras

Como afirmábamos, ninguno de los dos niveles anteriores descubren con exactitud lo que le gusta o no le gusta a la gente. El nivel de aspecto actúa en un marco de análisis mucho más preciso. En vez de fijarse en las construcciones del lenguaje (documentos, párrafos, oraciones), este nivel de análisis busca directamente a la opinión por sí misma. Para lograrlo, la asunción de este nivel se basa en la idea de que una opinión consiste en un sentimiento (positivo o negativo) y se refiere a algo en concreto (evento, suceso, producto, persona, etc.). (Liu, 2012)^[335]

La intención de este nivel es clasificar el sentimiento en relación a las características (aspectos) específicos de cada entidad. El primer paso consiste en identificar las entidades y sus características y considerar que los opinantes pueden dar diversas opiniones en función de cada aspecto de una misma entidad, como por ejemplo en esta frase: “La calidad de voz del teléfono no es muy buena, pero la duración de la batería es larga”. (Medhat, 2014)^[406]

5.4. Campos y posibilidades de aplicación del *Sentiment Analysis* (SA)

Como hemos afirmado en repetidas ocasiones, la opinión que la gente tiene de los negocios o de las organizaciones es muy importante para que estas puedan saber cómo se perciben sus productos o las acciones que llevan a cabo. Por las mismas razones, los gobiernos quieren saber qué piensa la opinión pública acerca de sus decisiones políticas. En función de las opiniones expresadas, el gobierno podrá tomar nuevas decisiones que permitan responder a los rápidos cambios sociales, económicos o políticos que han provocado sus acciones.

Una prueba firme de la notoriedad del SA en los entornos empresariales es que grandes corporaciones siguen desarrollando sus propios modelos para realizar los análisis, desde Google y Microsoft hasta Bloomberg y Amazon. En el nivel interno de los organismos políticos, las agencias también están monitorizando los medios sociales para hallar los intereses de la ciudadanía y saber aquello que piensan sobre determinados temas. (Liu, 2015)^[372]

Con el crecimiento de los medios sociales constatado en las dos primeras partes de esta tesis, ni las empresas ni otro tipo de entidades necesitan forzosamente realizar encuestas o *Focus Groups* para reconocer las opiniones de sus clientes o de sus seguidores. La información, como se ha mostrado, la encontrarán en gran parte disponible públicamente en la red. Del mismo modo, gracias a la monitorización de las redes sociales, podrán obtener información con contenido valioso y fiable. (Liu, 2015)^[372].

En este abanico creciente de campos de aplicación del SA entran también la escala individual y el propio interés de los ciudadanos y consumidores; hemos ido mostrando que los individuos también quieren saber lo que piensan otras personas de los productos, servicios o candidatos políticos antes de que ellos mismos deban tomar decisiones con respecto al mismo.

En los últimos años, las aplicaciones prácticas del SA han abarcado desde el estudio de productos hasta la sanidad, el turismo, los servicios financieros, los eventos sociales o las elecciones de gobierno, entre muchos otros campos.

Ofreceremos seguidamente casos que nos resultan especialmente singulares ya que han tenido a Twitter como fuente principal de obtención de datos.

5.4.1. Aplicaciones del SA en Twitter

En el ámbito de la investigación académica, distintos investigadores han aplicado el SA en las conversaciones de Twitter para predecir el éxito de una película y las ventas en taquilla (Mishne, 2006)^[418]. En concreto, Liu et al. aplicó el SA en las conversaciones de Twitter para predecir los ingresos que podía tener una película y descubrió que se podían predecir con relativa facilidad y precisión. (Liu, 2007)^[380]

Otro caso importante de SA en Twitter es que se han realizado modelos de regresión lineal capaces de predecir resultados electorales utilizando el volumen total de tuits (separados entre positivos y negativos) como variables independientes al mismo tiempo que los resultados de los sondeos de intención de voto (Bermingham, 2011)^[419]. Existen también trabajos que han utilizado la anotación manual de sentimientos en tuits con finalidades de estudio electoral. (Diakopoulos, 2010)^[420] (Sang, 2012)^[421]

En el campo de las finanzas y del funcionamiento bursátil se han realizado numerosas aplicaciones con intención predictiva (Das, 2001)^[373]. Se identificaron los mensajes del muro en una clasificación de tres tipos: optimista, pesimista o neutral. Los sentimientos resultantes respecto a cada uno de los valores bursátiles se agregaron y se utilizaron para predecir el *Mordan Stanley High Tech Index*.

Zhang et al. (2010) identificaron los estados de ánimo positivos y negativos en determinadas conversaciones de Twitter y los utilizaron para predecir el movimiento del valor de las acciones en índices como el Dow Jones, S&P 500 y el NASDAQ (Zhang, 2010)^[422]. Su investigación demostró que cuando las emociones en Twitter son muy agitadas, es decir, cuando muestran esperanza, miedo o preocupación, el Dow Jones baja el día siguiente. Cuando la gente expresa menos esperanza, miedo o preocupación, el Dow Jones sube el día después. Sus resultados señalaron que para lograr un modelo de predicción estándar se tienen que entrar a valorar sentimientos como calma y felicidad, pero que no todos serán útiles. Si et al (2013) combinaron los sentimientos expresados en series temporales con las fluctuaciones

bursátiles para predecir los movimientos diarios del índice S&P 100 mediante un Vector de Autoregresión. (Si, 2013)^[423]

Además de las finanzas, la política y las predicciones comerciales, que son tres de los campos más populares de estudio, se han realizado numerosas investigaciones en otra multitud de campos. Por ejemplo, las reseñas de productos se utilizaron para establecer un ranking de productos y comerciantes (McGlohon, 2010)^[424]. Se han estudiado las relaciones entre las apuestas *online* en la *National Football League*, las opiniones en blogs y las conversaciones en Twitter (Hong, 2010)^[425]. Tomando en consideración más redes sociales, se ha estudiado la fluctuación de las emociones a lo largo del tiempo y uso (Miller, 2011)^[426]. En otro caso, los sentimientos en los hombres se utilizaron para averiguar si se hallaba las diferencias de género respecto a los principales ejes emocionales. (Mohammad, 2011)^[427]

Como reflexiones finales del apartado que hemos dedicado al *Sentiment Analysis*, resaltaremos en primer lugar la relación directa que puede establecerse entre el rápido y eficaz desarrollo de sus métodos y aplicaciones y el crecimiento exponencial del contenido que los usuarios difunden públicamente en la web. En menos de una década se ha podido acceder en formato textual y con libertad de acceso a una cantidad antes inimaginable de opiniones de usuarios en las que los sentimientos, las emociones y sus juicios acerca de todo tipo de temas tienen una importancia capital. Es lógico que haya sido con la proliferación de este contenido disponible que, tanto desde el ámbito académico como desde el empresarial y político, hayan ido aumentando los recursos y el interés por extraer con su análisis información relevante.

Con el fin de situar a los investigadores y los trabajos realizados en este campo que consideramos de referencia, hemos presentado una serie de definiciones del SA, con aclaraciones respecto al concepto de *Opinion Mining*. Cada definición nos ha permitido sumar matices para obtener una visión completa a la hora de abordar su metodología. Antes de adentrarnos en la tipología de SA, hemos querido detenernos en la definición y explicación de los diversos conceptos que actualmente engloba el término inglés *sentiment* y que hemos convenido en traducir como “sentimiento”. Entendiéndolo desde esta perspectiva, “sentimiento” incorporará las opiniones y emociones como equivalentes en el análisis. Hemos señalado las razones que llevaron a los investigadores a tomar esta decisión.

En la descripción y el desarrollo de la tipología de SA hemos intentado realzar el notable cambio procedimental que conlleva adoptar una perspectiva *Machine Learning* o, por contra, una perspectiva basada en el léxico. Hemos apuntado la existencia de métodos que combinan una parte de ambos métodos.

En cuanto a las aplicaciones del SA, tanto en sus posibles campos como en ejemplos reales de investigaciones y estudios, nos ha interesado retomar los investigadores y académicos de referencia para esta tesis y mostrar, junto con otros breves ejemplos, casos en los que el SA se ha aplicado al contenido (tuits) compartido por los usuarios de Twitter. En algunas de las investigaciones citadas, se habían alcanzado previsiones basadas en los resultados del SA, así como detectado ciertos patrones de comportamiento de los usuarios. Estos ejemplos son pruebas excelentes que nos permiten proceder al tercer capítulo de la tesis con nuestra

particular aplicación del SA a las conversaciones de Twitter con el objetivo de poder hallar los *insights* publicitarios.

SEGUNDA PARTE

6.- SETAL (*SENTIMENT TEXT ANALYSIS*)

6. SETAL (SENTIMENT TEXT ANALYSIS)

En el capítulo anterior hemos descrito y analizado las técnicas y los algoritmos más utilizados por los investigadores en los ámbitos de la Teoría de Grafos, el RAT y el SA. Debido a la magnitud de los datos que manejan los estudios realizados dentro de cada uno de los ámbitos, los procesos de identificación y extracción de la información deseada, así como su posterior análisis, se llevan a cabo mediante el uso de algoritmos específicamente obtenidos para resolver cada una de las tareas.

Por el mismo motivo que exige el uso de sofisticados algoritmos, la enorme cantidad de texto también precisa herramientas con capacidad de análisis automático.

Nuestro objetivo en este capítulo es presentar la labor realizada por el Dr. Enric Monte, Ingeniero en Telecomunicaciones y profesor titular del departamento de Teoría de la Señal y Comunicaciones de la Universitat Politècnica de Catalunya y con quien hemos colaborado para el desarrollo del algoritmo de resumen y análisis de textos.

Al algoritmo lo hemos denominado *Sentiment Text Analysis* (SETAL), que no corresponde a ningún nombre comercial o marca registrada. Por ahora, SETAL no se ha desarrollado comercialmente y su nombre es genérico y descriptivo de su función.

La herramienta será capaz de desarrollar las siguientes funciones:

- Identificar los textos más relevantes en función del ámbito de interés.
- Analizar gramaticalmente los textos más relevantes.
- Calcular los sentimientos expresados en los textos analizados.

En esta tesis utilizaremos una fase beta de la herramienta que nos permitirá obtener la información básica necesaria para la elaboración de *insights* publicitarios.

6.1. Descripción de SETAL

El sistema SETAL es capaz de buscar y encontrar comentarios realizados por los usuarios de las redes sociales y elaborar con su contenido un resumen en función de unos criterios específicos o intereses concretos basados en palabras clave, definidas previamente en un diccionario (*lexicon-based*).

SETAL resume y jerarquiza los textos por orden de importancia en los tres niveles que hemos definido: palabras clave, oraciones y textos. Agrupa los contenidos por tópicos y define el perfil gramatical de cada usuario a partir del vocabulario que utiliza y extrae los sentimientos que han expresado a través de los textos.

El nuevo algoritmo en el que se basa opera desde varios ámbitos de forma simultánea. Por un lado, es capaz de identificar y capturar los textos de interés y descartar el resto. Por el otro, en el resumen que realiza aplica un filtro que elimina todo texto que no tiene o no puede aportar ningún valor para el análisis posterior. Finalmente, con el texto considerado como válido, realiza un análisis semántico capaz de identificar múltiples facetas del usuario y diversas variables del contenido (definidas anteriormente como la polaridad emocional y los sentimientos).

Dentro del RAT, la herramienta ordena las oraciones utilizando una puntuación al conjunto de las palabras clave que lo componen. La puntuación se elabora según la coocurrencia⁶⁸ de palabras clave, las propiedades de la Teoría de Grafos y su distribución estadística (parecida al mencionado Google *PageRank*, a partir de una cadena de Markov⁶⁹). La selección y resumen de textos se realizan a partir de la elaboración de un diccionario de palabras clave relacionadas con la temática de interés. Este diccionario se elabora ad hoc y está al servicio de los objetivos y del sector de actividad, producto o marca sobre el que se quiere realizar el estudio.

⁶⁸ Coocurrencia: la colocación de dos o más palabras que aparecen con cierta frecuencia en el discurso de manera que su búsqueda puede realizarse midiendo, entre otros aspectos formales, la frecuencia de cada uno de ellos, la distancia que tienen entre sí en un texto y el número de veces que se muestran juntos. [Sinclair, 1991]^[428]

⁶⁹ Cadena de Markov: define que la probabilidad de que un evento ocurra. Ello depende solamente del evento inmediatamente anterior y no de la secuencia de acontecimientos que la precedieron. [Markov, 1971]^[429]

Nuestra labor de investigación se ha centrado en el sector bancario/financiero español, argumentado en la sección cuarta: identificación de la muestra. Hemos desarrollado un diccionario sobre terminología específica del sector con un total de 816 palabras. Para la confección de este diccionario se ha recurrido al análisis de los productos financieros de los principales bancos españoles y del buscador Google, de tal modo que hemos comparado la lista propia con las palabras clave más buscadas por los usuarios en Internet (según el buscador Google) sobre este ámbito. (Ver Anexo: Diccionario sector bancario).

La herramienta ofrece el resumen de los textos en los siguientes formatos:

- Palabras más utilizadas
- Combinaciones de palabras más frecuentes
- Análisis gramatical detallado:
 - uso de verbos, adverbios, adjetivos,...
- Palabras más importantes (*)⁷⁰
- Oraciones más importantes
- Textos más importantes
- Tópicos:
 - grupos de palabras/oraciones que aparecen más frecuentemente.
- Usuarios más importantes:
 - Aquellos que tienen oraciones más importantes
- Clasificación de los Sentimientos por:
 - Tiempo, frase, usuario
- Sentimientos de los usuarios:
 - Clasificados por puntuación de sentimientos positivos / negativos.

Con este análisis se podrá obtener una “huella gramatical” que permitirá, a su vez, crear la “huella digital” de un usuario.

⁷⁰ Importancia según SETAL: el algoritmo de SETAL considera “importantes” aquellas palabras, oraciones, documentos o usuarios que obtienen una puntuación más alta en relación al tema de interés. Como hemos detallado en el capítulo anterior, esta puntuación no se obtiene por una mayor frecuencia, sino por su posición de centralidad en el grafo de textos y su distribución estadística.

En la primera fase de análisis, SETAL se basa en el modelo *lexicon-based*; utiliza el diccionario *Linguistic Inquiry and Word Count (LIWC)*, desarrollado por Pennebaker et al⁷¹ considerado el diccionario estándar para el cálculo de los sentimientos.

Con el cumplimiento del primer paso *lexicon-based*, seguidamente la herramienta incorpora métodos de aprendizaje automático ML y que analiza las palabras para identificar las que resultan más relevantes.

En nuestro caso, se tratará de puntuar una oración según el diccionario de palabras y/o conjunto de palabras que los grupos de investigación en psicología han determinado que están asociadas a una determinada intensidad de sentimiento/emoción.

Como explicamos, con esta metodología se le aplica a cada palabra un valor numérico que corresponde a la intensidad de los sentimientos/emociones y de su positividad o negatividad. En caso de no tener valor positivo ni negativo, la inexistencia de esta intensidad puede considerarse neutral.

La herramienta se ha desarrollado a partir del diccionario de Pennebaker et al. Y se le han aplicado mejoras considerando las especificidades del lenguaje castellano. También se ha enriquecido con el añadido de factores multiplicadores de la intensidad de los sentimientos según la organización gramatical, la presencia de adjetivos o adverbios y del tiempo gramatical de la oración, dando más importancia al tiempo presente que al pasado o futuro. (Ver Anexo: Diccionario LIWC).

⁷¹ Descripción detallada en la sección: Análisis de sentimientos, métodos

6.2. Metodología SETAL para la obtención de *insights*

Del mismo modo que analizábamos las técnicas propias que algunos anunciantes y agencias de investigación de mercados han desarrollado para la elaboración de *insights*, a continuación vamos a presentar una técnica propia que recoge las características específicas del medio Internet, las redes sociales y aglutina la tecnología necesaria para el análisis de grandes volúmenes de información. Con todo ello podremos realizar el proceso de extracción de la información de interés y eliminar el ruido existente para su posterior análisis.

Podríamos llegar a establecer cierta similitud entre la labor de investigar en Internet con la labor de los buscadores de oro. El mayor símil lo encontramos en esa ingente necesidad que ellos tenían de filtrar toneladas de tierra para obtener unos pocos gramos de oro. Nosotros necesitaremos filtrar “toneladas de texto” para hallar esos fragmentos de información realmente valiosos. Por otro lado, un *insight* es, para los anunciantes y las agencias de publicidad, como oro en bruto: gracias a los expertos en marketing y publicidad se acabará convirtiendo en una pieza de joyería con un alto valor comercial. Twitter comunicaba que 500.000.000 tuits se enviaban cada día a través de su sistema (Twitter, 2015)^[430]. Ese “Big Data” es como la mina de oro y el oro que encontremos será el éxito del proceso de extracción y análisis que llevaremos a cabo.

La metodología que detallaremos a continuación la podríamos dividir en cuatro grandes áreas:

- Fase de obtención de la información
 - Preproceso del texto original y preparación para su análisis
- Fase de análisis de la información
 - Algoritmo de decisión: frecuencias, posiciones, interacción y sentimientos
- Fase de resumen de la información
 - Resumen: realizar listas ordenadas y jerarquizadas por importancia
- Fase de elaboración del *insight*

6.3. Consideraciones del funcionamiento de SETAL

- Obtención de los datos:

Búsqueda de los comentarios (tuits) sobre un tema de interés realizada a partir de una selección previa de usuarios activos definidos como *influencers*.

- Usuarios *influencers* más relevantes:

Gracias al diccionario creado para la ocasión, SETAL identifica y jerarquiza aquellos usuarios que más comentarios han realizado sobre el tema de interés definido y que tienen más de 5.000 seguidores.

- Filtro de contenidos:

Seleccionados los usuarios más relevantes, se filtran los tuits para eliminar el “ruido” y los tuits irrelevantes para el objeto de estudio⁷². Análisis gramatical:

Determinará la frecuencia de uso de las partículas gramaticales que hemos especificado.

- Cálculo de los sentimientos:

Con el diccionario LIWC como base, SETAL puntúa cada oración según sus palabras y los grupos que forman. A cada tuit se le atribuye una puntuación que corresponde a la intensidad de sentimientos de positividad/negatividad.

- Jerarquización:

El conjunto total de tuits se ordena a partir de la puntuación obtenida. Esta puntuación vendrá dada por la coocurrencia de las palabras, las propiedades de la Teoría de Grafos y su distribución estadística.

Formación de una lista ordenada con las oraciones y sus sentimientos asociados de mayor a menor importancia. Se obtienen los conceptos y expresiones más importantes expresadas por cada uno de los usuarios.

⁷² Criterios detallados en la sección cuarta: identificación de la muestra

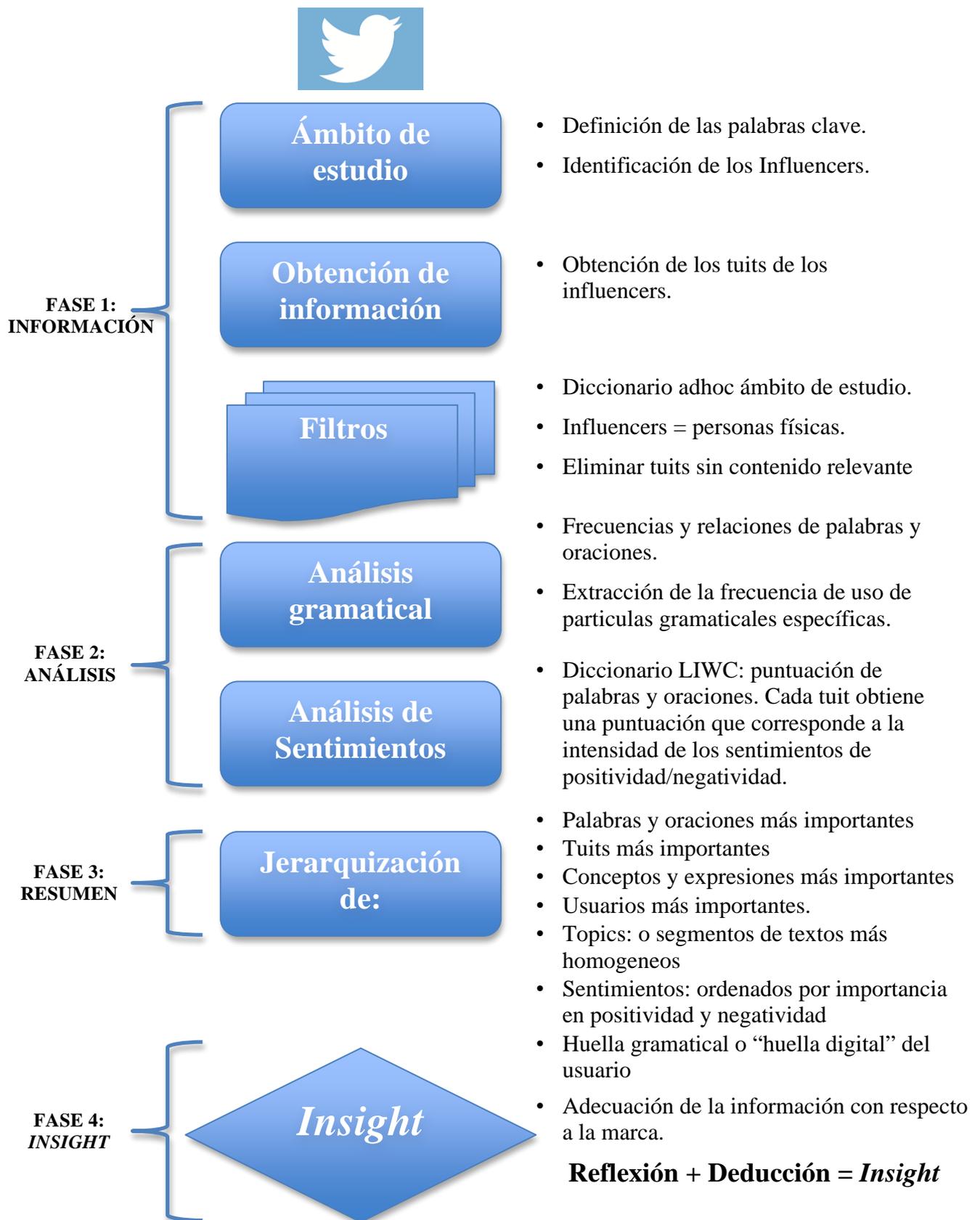
Obtención de expresiones breves y concisas de los sentimientos gracias a los usuarios y a las puntuaciones que han recibido sus expresiones.

Establecimiento de la huella gramatical de los usuarios con la información que los caracteriza. Consideramos que esta huella incorpora una información muy relevante para su mejor conocimiento.

- Relevancia de los resultados:

Recordamos aquí la importancia de haber obtenido toda esta información sin nuestra intervención ni participación directa en los comentarios. Éstos se han realizado de forma libre y espontánea por parte de los mismos usuarios. Hemos conseguido reducir al mínimo el sesgo producido durante los procesos tradicionales de investigación.

Siguiendo el esquema a continuación, detallamos y describimos las diversas fases a tener en cuenta (metodología y esquema propios):



6.3.1. Fase de obtención de la información

6.3.1.1. Ámbito de interés

Dentro de la comunicación publicitaria, un ejemplo de ámbito de interés puede ser el estudio de una marca o producto en el mercado y de su competencia directa.

En este caso, la acotación de la red social Twitter empezará por identificar aquella información que contenga las palabras clave relacionadas con el producto o servicio objeto de nuestro estudio.

Como criterio de elaboración de la lista de palabras clave hemos elegido aquellas que mejor definen el producto o servicio y las hemos comparado con las búsquedas que los usuarios han realizado en el buscador Google. Una vez obtenidas ambas listas, las hemos comparado manualmente y ordenado según su volumen y adecuación a las búsquedas.

Con la lista definitiva de palabras clave del ámbito de interés hemos procedido a identificar a los *influencers* dentro de nuestro ámbito.

6.3.1.2. Identificación de los *influencers*

El criterio de selección de los *influencers* consiste en identificar a aquellas cuentas de Twitter (con 5.000 seguidores o más) que correspondan a usuarios particulares (eliminando a las cuentas corporativas) y que en sus tuits haya aparecido alguna de las palabras clave seleccionadas. Para evitar sesgos, no contemplamos el estudio de usuarios de Twitter que son seguidores de la marca. La finalidad de todo ello es proponer una metodología lo más aséptica, neutral y transparente.

6.3.1.3. Obtención de la información

Con la lista de palabras clave y los *influencers* identificados, se procede a guardar todos los tuits enviados por los *influencers* durante el período necesario para obtener una muestra óptima y estadísticamente.

La cantidad variará en función de los objetivos, pero tendremos en cuenta que casi el 99% del contenido se podrá considerar “ruido”. La cantidad resultante, después de realizar los filtros adecuados, debe tener un nivel de representatividad superior al 0,007% (equivalente al nivel de representatividad de 2.000 encuestas sobre el total de la población) que obtendría un nivel de fiabilidad probabilística de $\pm 2,2\%$.

Como ya hemos comentado y profundizaremos en la próxima sección, la cantidad de tuits guardados durante tres meses fue de 1.200.000.

6.3.1.4. Filtros: eliminación del ruido

Resulta evidente que para la elaboración de *insights* se requieren textos con contenido relevante para la marca. Como veremos en el siguiente capítulo,⁷³ Twitter ofrece muchas ventajas para el análisis de sus contenidos, aunque también tiene inconvenientes fruto de sus propias características (máximo de 140 caracteres por tuit, alto uso de abreviaciones, emoticonos, argot e ironía). Los filtros serán imprescindibles para poder separar el “ruido” de los textos realmente relevantes.

Los criterios utilizados para la elaboración de los filtros pueden variar según las necesidades. Los filtros establecidos y propuestos para nuestra metodología los encontraremos en el capítulo cuatro que destinamos a su explicación. Llegados a este punto, procederemos al análisis.

⁷³ Capítulo 4, fragmento perteneciente al punto “Identificación de la muestra”.

6.3.2. Fase de análisis de la información

6.3.2.1. Análisis gramatical de los tuits

El análisis gramatical nos permite obtener información relevante sobre el contenido tanto semántico como gramatical de los textos seleccionados.

Con los criterios ya señalados, el análisis gramatical identifica y jerarquiza las palabras más importantes. Aplicando la Teoría de grafos, consideraremos las palabras como nodos conectados entre sí. Su posición y sus conexiones (enlaces) dentro del conjunto de todas las palabras le conferirán una mayor o menor importancia dentro del conjunto total de los textos analizados y, finalmente, obtendremos su distribución estadística.

El análisis gramatical se repetirá en los otros niveles de análisis: oraciones, tuits, documentos y el conjunto de todos los textos analizados.

La información obtenida se separará por variables de análisis: palabras, relaciones entre palabras más frecuentes, verbos, adverbios y también según el uso de pronombres.

Del mismo modo que la técnica desarrollada por Nestlé hablaba de “palabras frecuentemente utilizadas”, “alto recuento de palabras en investigación”, SETAL también realiza un recuento de la frecuencia con el estudio del concepto definido anteriormente como “centralidad”.

6.3.2.2. Análisis de los Sentimientos (*Sentiment Analysis*)

El análisis de los sentimientos (*sentiments*) es el complemento imprescindible al análisis gramatical mediante la incorporación de la polaridad específica de los términos y unidades de análisis elegidas.

Con el diccionario LIWC como referencia, se inicia un proceso de puntuación numérica a nivel de palabras, oraciones y tuits que permite la jerarquización según la polaridad de la carga emocional.

Si Buscamos similitudes con otras técnicas ya desarrolladas para la obtención de *insights*, Nestlé usaba el concepto de “palabras mágicas” para definir aquellas que además de su alta frecuencia expresaban “convicción o emoción”, ya fuesen positivas o negativas”. Como veremos, nuestra herramienta SETAL también completa las variables cuantitativas con las cualitativas de forma automática.

6.3.3. Fase de resumen de la información

6.3.3.1. Jerarquización y resumen de toda la información

En la fase anterior se obtenían las métricas tanto cuantitativas como cualitativas de todos los textos analizados. En esta tercera fase, SETAL nos presenta la información de forma ordenada jerárquicamente y según sus diversas características. La importancia de las listas no residirá exclusivamente en la frecuencia de aparición de los términos sino en el papel que una palabra u oración juega dentro del conjunto de todos los textos analizados.

La ordenación de la información nos presenta los datos ordenados según los contenidos de interés siguientes:

- Palabras más importantes: jerarquiza las palabras más frecuentes y centrales que aparecen en los textos. Se establece un vocabulario por orden de importancia.
- Oraciones más relevantes: parte del mismo criterio anterior pero lo sitúa al nivel de las oraciones.
- Tuits más remarcables: Sigue el mismo criterio que el inicial pero aplicado al tuit completo. En ocasiones, el tuit más importante puede coincidir con la oración más importante, en función de su longitud.

- Conceptos y expresiones más importantes: variación específica del mismo criterio: estos conceptos y expresiones pueden no corresponder ni a oraciones ni a tuits, sino a sus partes.
- *Topics*: los *topics* son segmentos de textos más homogéneos, comunes en diferentes usuarios, que permiten segmentar a dichos usuarios por la similitud en el contenido de sus comentarios.
- Polaridad y puntuación de los sentimientos: jerarquización de las oraciones, tuits y usuarios en función de la puntuación obtenida por la expresión de sus sentimientos.
- Usuarios más importantes: se determinan según la importancia que han tenido sus tuits. Se obtiene una lista que jerarquiza los usuarios por orden de importancia relacionada por los comentarios más cercanos a nuestro ámbito de interés.
- Sentimientos por períodos: los sentimientos analizados en los tuits se pueden asociar a determinadas horas, días o meses. Ello permite identificar las horas, días y semanas en que los usuarios tienden a expresar sentimientos más positivos o negativos.
- Huella digital: se puede analizar cada usuario de forma individual, de tal modo que nos permite obtener una idea de algunas de sus características personales a partir de su forma de expresarse en Twitter.

En esta fase, los investigadores han obtenido una información amplia y detallada del objeto de estudio y de los usuarios seleccionados. Tienen información sobre las palabras más importantes aparecidas dentro de los tuits, las oraciones y las expresiones. También se ha obtenido información cuantificada y jerarquizada sobre los sentimientos expresados por los usuarios (incluyendo su separación por horas, días y meses). Se habrá podido elaborar la huella digital, con información precisa acerca de cada usuario, así como de los principales *topics*, ya sea a nivel concreto o de manera global. En este sentido, la obtención de un *insight* podrá lograrse tanto desde una perspectiva de análisis global del grupo como analizando a un usuario concreto o una de sus oraciones o expresiones emitidas.

El proceso de análisis realizado hasta el momento nos ha permitido alcanzar ya un alto conocimiento de los usuarios: cómo son, qué les gusta, qué quieren, qué sentimientos tienen

con respecto a un producto/marca, qué puede ser esencial o superficial para ellos, cómo se divierten, etcétera. Nos sirve recordar aquí el método que presentaba Unilever para la obtención de *insights*.

El método propuesto en esta tesis también realiza una inmersión en la observación del consumidor, tal y como indica el método de Salvetti-Llompарт. Como dicho método indicaba, es imperante observar, descubrir y comprender las necesidades, motivaciones, deseos, los instintos y las diversas identidades de los usuarios para luego ubicarlos en un mercado concreto, en una situación determinada o entender qué uso darían de un producto o servicio elegido. El método que hemos propuesto permite lograrlo evitando, como hemos ido asentando a lo largo de la tesis, el uso de métodos tradicionales que implican la intervención directa de los investigadores.

6.3.4. Fase de elaboración del *insight*

La fase de elaboración del *insight* es la fase del descubrimiento fruto de la asimilación, la reflexión e incluso de la expresión intuitiva de toda la información procesada y analizada previamente.

Nestlé lo considera el momento del “¡Ajá!, aunque como sabemos es un descubrimiento que llega después de un largo y detallado proceso de estudio de los usuarios y de toda la información relacionada.

Para hallar los *insights*, Unilever agrupa toda la información en diferentes plataformas (consumo, públicos objetivos, motivos de compra, tendencias, entre otros).

Salvetti-Llompарт apuesta por la intuición surgida gracias a toda la información obtenida y de obtener los *insights* según los mercados, los productos y las situaciones.

Finalmente, Quid separa los *insights* entre los del “consumidor” y los de “marca”. Los del consumidor apelan a sus vivencias, actitudes, motivaciones, miedos y deseos. Los de marca son la adaptación de los *insights* del consumidor a la realidad de la marca en concreto. Una vez fijados los objetivos y extraídos los *insights* del consumidor, el tercer paso será aplicarlos a los *insight* de la realidad de la marca.

En esta tesis establecemos como criterio esencial para el éxito en la obtención de los *insights* el haber alcanzado previamente un conocimiento profundo tanto del consumidor como de la marca.

Para lograrlo, será importante que los investigadores focalicen su atención en aquellas informaciones que son realmente relevantes, prestando especial interés en los resultados jerarquizados y segmentados que ha ofrecido el análisis automático.

Con la aplicación del SA y de nuestra herramienta SETAL, apostamos por la obtención de análisis de carácter empírico y que nos pueden proporcionar *insights* que dependen menos de la “inspiración” y la “espontaneidad” que del verdadero trabajo científico. El llamado momento “¡ajá!” no lo contemplamos como una opción a tener en cuenta de manera sistemática. Aunque la obtención de los *insight* precisa creatividad e ingenio, gracias a la aplicación de los métodos científicos como el que acabamos de presentar evitamos muchísima pérdida de tiempo e información y ganas en eficiencia, seguridad y precisión.

TERCERA PARTE

7. INVESTIGACIÓN EMPÍRICA

7. INVESTIGACIÓN EMPÍRICA

En las páginas que siguen a continuación se presentará la metodología de la investigación empírica que se ha desarrollado para esta tesis. En el capítulo quinto se detallará la metodología utilizada para la parte de la investigación cualitativa.

La investigación cuantitativa se ha desarrollado mediante la técnica del análisis de contenido, es decir, hemos realizado un análisis de las frecuencias de aparición de una serie de palabras clave presentes en los tuits de una selección de usuarios a partir de un diccionario desarrollado para este estudio. Posteriormente, la herramienta ha seguido analizando los contenidos de los textos desde diferentes puntos de análisis (semántico, gramatical y de los sentimientos expresados).

El estudio del análisis lo hemos realizado con la técnica de la netnografía. La fuente principal de obtención de la información ha sido la observación de un grupo de usuarios activos en Twitter. En la posterior sección “Identificación de la muestra” explicamos los motivos de la elección de los usuarios y del por qué de la elección de Twitter.

El seguimiento de los usuarios ha sido de tres meses (las variables cuantitativas se especifican en el punto 4.3), durante los que hemos obtenido información sobre su actividad diaria (qué dicen, cómo lo dicen) y en especial acerca de cómo son sus actitudes, emociones y sentimientos en relación a diferentes aspectos de la vida cotidiana. Ha sido fundamental la elección correcta de los *influencers*, puesto que de su elección dependerá en gran medida la obtención de unos u otros resultados.

Para realizar el análisis de los contenidos de los usuarios hemos tenido en consideración dos técnicas descritas anteriormente. Por un lado, la Recuperación de información (*Information Retrieval* o IR) y, por el otro, el aprendizaje automático (ML). Recordemos que el IR nos filtrará los textos partiendo de una lista de palabras clave y el ML lo considera una clasificación estándar.

La diferencia entre ambas técnicas reside en la forma o fuente de alimentación de la información requerida. En el IR se genera una lista de palabras clave relacionadas con el

objeto de estudio, como por ejemplo una marca, mientras que en el segundo caso se fija como fuente de información determinada una selección de fuentes adecuadas (como Twitter o los blogs, por ejemplo), que servirán como “entrenamiento” o base de ejemplo sobre la que crear un modelo que servirá para desarrollar posteriormente la investigación.

Según los resultados que se vayan obteniendo, se necesitará realizar ciertos ajustes mediante un proceso de adecuación manual. Por un lado, deberá ir ajustándose el número de palabras clave y, por el otro, la selección de las fuentes más representativas del sector. Siendo por un lado la adecuada selección de palabras clave como por el otro.

En cuanto a la selección de la información realizada por SETAL, existen algunas variables tenidas en cuenta dentro del ámbito de la Teoría de grafos y presentes en el algoritmo: el flujo de intermediación (*betweenness*), el criterio de cercanía (*closeness*) y el grado de centralidad (*degree centrality*). Todas ellas son medidas de autoridad e influencia (tanto de un usuario y de los textos) más precisas y sofisticadas que el número total de repeticiones de una palabra o la cantidad total de seguidores de un *influencer*.

En este sentido, Takashi ha propuesto un sistema de modelado del lenguaje para la extracción automática de contenidos a partir de dos factores: *phraseness* y *informativeness*. El primero determina el nivel o grado en que una secuencia de palabras se puede llegar a considerar una frase, midiendo la colocación y cohesión existente en las palabras que se organizan de forma consecutiva. *Informativeness* se refiere al nivel o grado en que una frase mejor ilustra las ideas clave o más centrales de una discusión (Takashi, 2003)^[431]. En su aplicación, el *phraseness* y el *informativeness* de las frases clave pueden calcularse utilizando la divergencia de Kullback-Leibler⁷⁴.

Con los mencionados *topics* hemos podido realizar otra aproximación a la identificación de informaciones relevantes gracias a la detección de grupos de mensajes representativos que hablan del mismo tema en un corto intervalo de tiempo. (Sharifi, 2010)^[368]

⁷⁴ Creada por Solomon Kullback y Richard Leibler en 1951. La KLIC es una medida no simétrica de la similitud o diferencia entre dos funciones de distribución de probabilidad P y Q .

Otras fuentes destacadas de investigación sobre palabras o frases clave son las conocidas como *Clustering Document*, *Latent Dirichlet Allocation* o *Probabilistic Latent Semantic Analysis*. Sirven también como herramientas de predicción.

El diccionario utilizado ha sido el LIWC2007, adaptado al catalán y al castellano, tal y como desarrollaremos en la metodología.

El paso adicional y objetivo de esta tesis es la correcta selección de las conversaciones más relevantes e importantes para una marca. Para poder obtener información útil para la identificación de *insights* e identificar, de forma automática los sentimientos, sean positivos como negativos. Esta información es de gran importancia para poder desarrollar acciones de marketing, que pueden afectar el producto, servicio, acciones de comunicación como el servicio de atención al cliente e incluso las propias campañas de comunicación publicitaria.

En este sentido se ha desarrollado una labor de investigación dentro del campo del Procesamiento del Lenguaje Natural PLN para detectar la polaridad de sentimientos expresados en las conversaciones de los usuarios. (Ver Anexo: Diccionario LIWC).

7.1. Consideraciones

7.1.1. Respecto a la metodología

Es este apartado apuntamos algunas consideraciones a tener en cuenta respecto a la metodología, técnicas y herramientas del desarrollo del estudio cuantitativo, explicadas anteriormente en los capítulos de desarrollo teórico.

La base metodológica fundamental para la cuantificación del proceso de investigación se ha centrado en el análisis de contenido, herramienta de investigación que permite analizar los contenidos de una forma no intrusiva que proporciona un método objetivo, sistemático y cuantitativo (Yale, 1998)^[432]. Nuestra metodología de investigación de análisis de contenido se centrará en el análisis de los discursos textuales, tal y como explicamos a través de Kozinets y Del Fresno (Kozinets, 2002)^[30], (Del Fresno, 2011:64)^[3] en los capítulos teóricos.

El análisis que desarrollaremos se centrará en la observación, monitorización y análisis de las cuentas de usuarios activos de la red social Twitter con el fin de obtener los contenidos más relevantes relacionados con los temas de nuestro ámbito de interés, siguiendo el procedimiento de investigadores como (Haas, 2004)^[432], (Langer, 2005)^[433] o (Nelson, 2005)^[5].

Hemos considerado oportuno no capturar ni analizar información acerca de temas de actualidad como la política, la vida de los famosos ni acontecimientos mediáticos como desastres naturales, accidentes o atentados terroristas, aunque es inevitable que hayan aparecido tuits relacionados.

7.1.2. Respecto al uso de la netnografía

Esta técnica se considera una potente herramienta de los profesionales del marketing y vendedores para desarrollar una profunda visión estratégica, implementar nuevas ideas de

innovación y explorar otros enfoques en función de la marca, las campañas de publicidad y la gestión de las redes sociales. (Kozinets, 2010)^[6]

Como anotamos, representa una modificación del término “etnografía”, convirtiéndolo en lo que llamaríamos “etnografía virtual” (Hine, 2000)^[434] o “etnografía online” (Markham, 2005)^[435] un neologismo que se adapta al trabajo de campo desarrollado en entornos de internet.

Por las razones expuestas en la definición de netnografía del capítulo segundo, que cuestionan la objetividad del estudio en el momento en el que se detecta la presencia del investigador, hemos preferido no revelar nuestra identidad como investigadores, como tampoco lo han hecho en otros casos anteriores investigadores como Langer, Nelson o Quinton (Langer, 2005)^[433], (Nelson, 2005)^[5] o (Quinton, 2010)^[436].

7.1.3. Respecto a la privacidad de los contenidos estudiados

Tomaremos algunas referencias respecto a la tipología y fuente de los contenidos estudiados y su relación con la pertinencia o no de que sean objeto de nuestro estudio. La conclusión a la que llegamos con la información encontrada es que no requerimos permiso para realizar nuestra investigación ni incurrimos en el estudio de información privada o confidencial.

En este sentido, Langer y Beckman argumentan que, del mismo modo que no se pregunta o informa a los autores de la bibliografía en la que se cita su labor, tampoco se considera necesario pedir permiso por citar contenidos publicados en sitios web o redes sociales cuyos contenidos son públicos y visibles por todo el mundo. (Langer, 2005:197)^[433]

Para Kozinets, la resolución de este problema varía en función si los contenidos analizados proceden de fuentes de información públicas o privadas (Kozinets, 2002:65)^[30]. Según Paccagnella, las conversaciones que se desarrollan en espacios digitales públicos se convierten en: “public acts deliberately intended for public consumption” (Paccagnella, 1997:6)^[437]. En consecuencia, no se pueden comparar con contenidos privados como los que se comparten entre usuarios a través de un correo electrónico o una carta ordinaria.

En nuestro caso, confirmamos que toda la información que hemos extraído corresponde a la que públicamente han compartido los usuarios elegidos. Nos hemos limitado a obtener y analizar sus tuits y hemos sustituido el campo del nombre (a menudo, seudónimos) por una cifra, sin indagar más allá en los datos personales del usuario. Toda la información obtenida para esta investigación procede de las publicaciones públicas de las cuentas de Twitter.

7.1.4. Respecto al uso de la folcsonomía

Respecto a la folcsonomía, siguiendo el esquema presentado en el capítulo segundo, consideramos los siguientes aspectos:

Como apuntamos previamente, el concepto de Thomas Van de Wal es el resultado del etiquetado realizado de forma libre por parte de los usuarios, de información y objetos (cualquier cosa con una URL) para su propio uso. El etiquetado se desarrolla en un entorno social (normalmente compartido y abierto a los demás). La folcsonomía se crea, por lo tanto, por el acto de etiquetado realizado por el usuario que consume información.

The value in this external tagging is derived from people using their own vocabulary and adding explicit meaning, which may come from inferred understanding of the information/object. People are not so much categorizing, as providing a means to connect items (placing hooks) to provide their meaning in their own understanding. (Van der Wal, 2007)^[9]

De esta definición destacamos el etiquetado de forma “libre y personal” en un “entorno social” y que “añade un significado explícito”. Gracias a la folcsonomía, los propios usuarios nos revelarán una serie de preferencias, opiniones y prioridades.

Señalamos a continuación las características de nuestro uso de la folcsonomía:

- Entorno: Internet
- Redes Sociales: Twitter
- Los usuarios: usuarios identificados y activos en Twitter
- Las etiquetas: palabras clave
- Los recursos: páginas web, fotos, vídeos y podcasts

- Otras folcsonomías tenidas en cuenta: emoticonos, Hashtags (#), tuits preferidos y dominios de sitios web elegidos i/o seguidos.

Nos ha sido de gran ayuda y utilidad el hecho de aprovechar que la función de las etiquetas en un estudio de SA como el nuestro es la misma que los usuarios les dan a las palabras clave, hashtags o emoticonos: gestionar, categorizar y resumir el contenido online acerca del que están hablando.

7.1.5. Respecto al análisis gramatical de los contenidos

Dentro de las herramientas de SETAL que hemos presentado, mencionamos las posibilidades que ofrece respecto a la detección de las “palabras funcionales” (*Function Words*) y las *palabras de contenido (content words)*.

7.1.5.1. Palabras funcionales (*Function Words*)

Aunque contengan un significado léxico pobre y a veces amiguo, las palabras funcionales sirven para establecer relaciones gramaticales con otras palabras dentro de una oración y también pueden describir el estado de humor y/o actitud del individuo. Su estudio es fundamental para proceder al análisis gramatical de las oraciones. <Mohr, 1994> Las palabras funcionales suelen ser, principalmente, verbos auxiliares, preposiciones, conjunciones y pronombres.

Hemos querido tener en cuenta las palabras funcionales basándonos en los descubrimientos de Pennebaker. El investigador mostró, por ejemplo, que los pronombres son singularmente relevantes para la clara determinación de algunos rasgos de la personalidad. (Pennebaker, 2003)^[438]

Pennebaker realizó estudios de transcripciones entre militares y descubrió su rango militar a partir de los patrones de su discurso; fueron las palabras funcionales y no las de contenido las que permitieron los hallazgos. En otros casos puede detectarse el género de la persona mediante el estudio del uso de ciertas palabras. (Fast, 2008)^[439]

Según Pennebaker, existen unas 500 palabras funcionales y 150 de ellas son de uso muy frecuente. Como apuntábamos, el mismo investigador afirmaba que su estudio nos permite tener una idea de su estado emocional, personalidad, edad, género e, incluso, de su clase social. (Pennebaker, 2011)^[440]

Avanzamos aquí algunos de sus hallazgos de forma concreta:

Los pronombres indican dónde las personas fijan su atención. El uso del pronombre “yo” muestra un foco muy marcado en uno mismo. Si alguien pregunta “¿Qué tiempo hace?”, las respuestas pueden ser tanto “hace calor” como “yo creo que hace calor”. Aunque el añadido del “yo creo” pueda parecer un detalle, en realidad está lleno de significado.

En el mismo sentido, se puede identificar si una persona está mintiendo porque tiende a usar “nosotros” más a menudo y deja de utilizar por completo frases con pronombres en primera persona. Por ejemplo, un mentiroso, en lugar de decir “yo no he cogido el libro” diría “esta no es la clase de cosas que haría una persona con integridad”. Las personas honestas usan “pero”, “sin”, “no”, “nadie” y “nunca” con mucha mayor frecuencia. Una persona depresiva utilizará el “yo” en un 6,5% de sus palabras mientras que una persona no depresiva lo haría en un 4%. Solo a través de análisis computerizados se pueden detectar estas diferencias, por muy significativas que sean. El sexo femenino utiliza “yo”, “me/mi” y “mío” más a menudo que el masculino. Por su lado, los hombres utilizan con mayor frecuencia los artículos “un/o/a” y “el/la”. Ello indica que los hombres hablan más acerca de objetos y cosas. Las mujeres utilizan más pronombres en tercera persona (“el”, “ella” y “ellos/as”) ya que hablan más a menudo de personas y amistades. (Pennebaker, 2011)^[440]

7.1.5.2. Palabras de contenido (*Content words*)

Las palabras de contenido poseen un significado completo, siendo en su mayoría nombres, sustantivos, verbos, adjetivos y la gran mayoría de adverbios.

Las palabras de contenido contrastan con las funcionales porque son de “clase abierta” y, por lo tanto, pueden ser añadidas al léxico. (Giménez de la Peña, 2003)^[441]

7.1.6. Respecto al uso del diccionario *Linguistic Inquiry and Word Count (LIWC)*

Como programa de análisis de texto, el LIWC es capaz de calcular cómo las personas usan diferentes categorías de palabras a través de una gran diversidad de textos. Es el programa que nos ha permitido identificar el grado en que los usuarios han utilizado palabras con connotaciones emocionales (positivas o negativas), así como diversas categorías de palabras. Su diccionario incorporado define qué palabras deben ser tenidas en cuenta dentro de nuestros archivos.

Los archivos de texto analizados pueden contener lo siguiente:

- 17 dimensiones lingüísticas estandarizadas (por ejemplo: conteo de palabras, porcentaje de pronombres o artículos)
- 25 categorías de palabras sobre aspectos psicológicos (por ejemplo: afecto o cognición).
- 10 dimensiones relacionadas con “relatividad” (por ejemplo: tiempo, espacio o movimiento)
- 19 categorías respecto a cuestiones personales (por ejemplo: trabajo, hogar u ocio).

La selección de palabras que definen las categorías de LIWC se ha desarrollado durante varios años. La idea inicial fue identificar un grupo de palabras que englobaban las emociones básicas y dimensiones cognitivas que han sido típicamente estudiadas en

psicología social, el área de salud y de personalidad. A partir de aquí, las categorías iniciales se han ido expandiendo.

A partir del diccionario original LIWC hemos desarrollado un diccionario adhoc en castellano y catalán, asegurando así que no quedaban palabras o expresiones específicas de ninguno de los dos idiomas sin incluir. Ello nos ha permitido precisar más el análisis de los tuits.

7.2. Identificación de la muestra

Antes de adentrarnos en cómo hemos trabajado con la muestra de la red social elegida, Twitter, realizamos un estudio previo de las tres redes sociales más utilizadas en España además de Twitter y explicamos el por qué de su no elección en esta tesis.

Realizamos la identificación de las redes sociales más utilizadas por los usuarios en España mediante la quinta edición del prestigioso Estudio anual de Redes Sociales elaborado por el Interactive Advertising Bureau en abril de 2014.

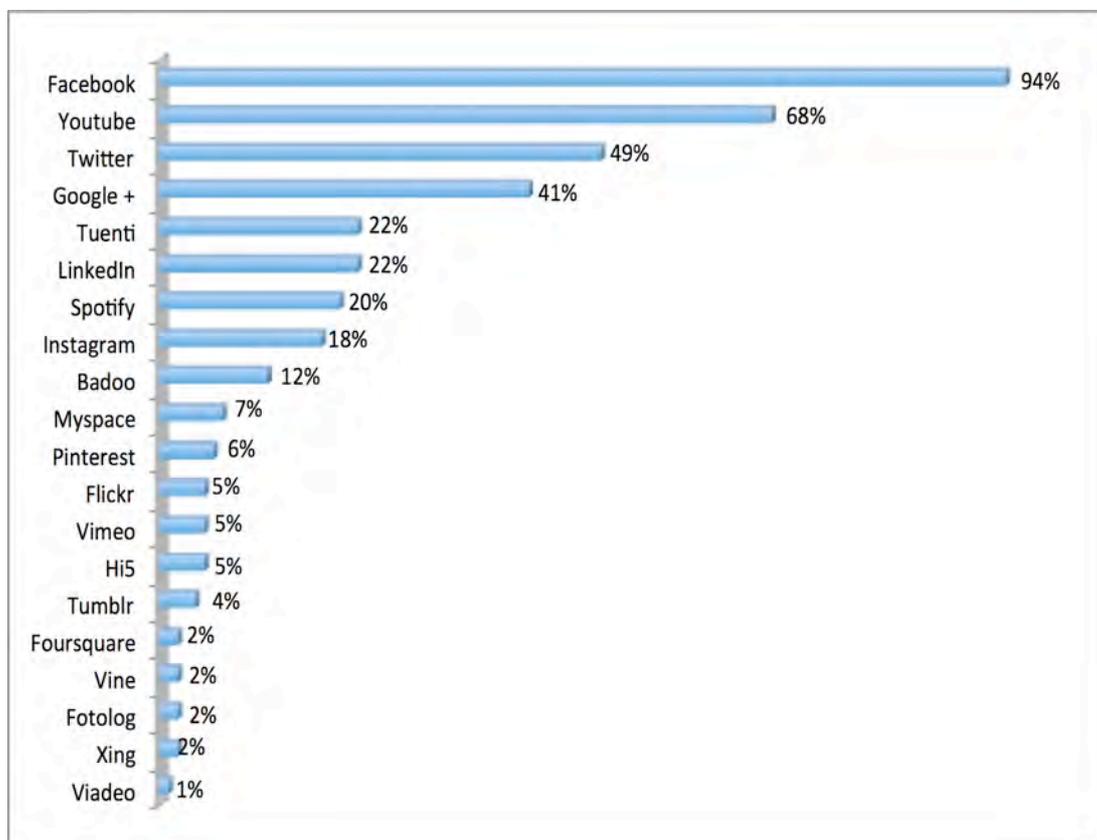


Figura 48: Gráfico de las redes sociales más utilizadas/visitadas en España. (Interactive Advertising Bureau Spain, 2015)^[442]

7.2.1. Selección de la red social

7.2.1.1. Análisis de Facebook

A priori, sus características son óptimas para identificar los sentimientos de sus usuarios, pero para nuestro caso presentaba dos problemas: el uso habitual y mezclado de todo tipo de formatos frecuentemente unidos y relacionados (vídeo, imagen y texto) y el problema de la privacidad, ya que una gran mayoría de usuarios lo utilizan sin permitir que la información que comparten se difunda de forma pública.

Si bien es cierto que hay investigadores que han considerado que Facebook es una red social óptima para el estudio de los sentimientos expresados por sus usuarios (Ahkter, 2010)^[443], también hemos encontrado una mayoría de investigadores que prefieren Twitter y que citaremos posteriormente.

7.2.1.2. YouTube

La razón principal por la que descartamos Youtube es por su enfoque mayormente dirigido al formato audiovisual. Los comentarios textuales que se encuentran en algunos vídeos no son suficientes para proceder con nuestra investigación.

7.2.1.3. Google Plus

Aunque esté situada como una de las más utilizadas, ello se debe a que Google obligó en su momento a utilizarla de manera obligada con el uso de otros servicios mucho más utilizados como Gmail. Actualmente presenta un uso muy bajo tanto a nivel de usuarios como de cantidad de publicaciones rastreables y los análisis realizados no nos ofrecían información suficiente como para emprender nuestra investigación.

7.2.1.4. Twitter

Twitter nos presenta diversas ventajas determinantes en cuanto a su marco general de uso: no presenta restricciones sobre el contenido que pueden difundir los usuarios (salvo que

vulneren la ley) y su contenido es mayormente público (aunque los tuits pueden ser “protegidos”, la propia dinámica y funcionamiento de la red lleva a la mayoría de usuarios a tener sus perfiles abiertos, en concreto, un 92% de los usuarios). Todo ello ha permitido que tanto la academia como la industria hayan realizado importantes investigaciones teniendo Twitter como fuente de información (Cha, 2010:12)^[444] y se ha utilizado en algunos casos por conocidos investigadores en el campo del SA. Finalmente, su crecimiento ha sido sólido, rápido y mundial en los últimos años. (Pang, 2008)^[331] (Sharifi, 2010)^[368] (Liu, 2012)^[335] (Si, 2013)^[423]

Sin embargo, como hemos apuntado con anterioridad, presenta también importantes limitaciones: 140 caracteres por tuit, las numerosas abreviaciones, el uso de emoticonos, una gran cantidad de irregularidades y, como culminación de todo ello, sobre todo mucho “ruido”⁷⁵.

Resumimos a continuación y de forma esquemática las principales ventajas de Twitter:

- Medio masivo.
- Utilizado de forma habitual como medio de comunicación interpersonal y entre marcas y clientes.
- Gran capacidad de acceso a una notable cantidad de información.
- Fácil acceso a la información (no requiere un desarrollo informático complejo).
- La propia plataforma ofrece herramientas que facilitan su gestión.
- Facilidad de uso.
- Uso gratuito.
- Finalidad mayormente informativa.
- Se utiliza fundamentalmente texto que facilita su análisis.
- La información que se publica en Twitter es accesible de forma inmediata y los algoritmos la pueden analizar de forma igualmente rápida.

Detallamos las siguientes ventajas adicionales que se derivan de las que acabamos de presentar:

⁷⁵ Clasificamos como “ruido” toda aquella información irrelevante para su análisis.

- Twitter ofrece la posibilidad de acceder directamente a la fuente de emisión de los tuits y ello nos genera con facilidad una gran cantidad de información disponible para ser analizada.
- Si bien los 140 caracteres de un tuit son poca información, la recolección consecutiva de todos los tuits de un usuario (o miles de ellos) nos ofrecen cientos de miles o millones de tuits. Además, el límite de caracteres lleva a que muchos usuarios lo utilicen de forma concisa, directa y precisa, yendo al núcleo del tema que quieren tratar y sin entrar en comentarios tangenciales o secundarios. (Liu, 2012:10)^[335]
- Mientras que un motor de búsqueda como Twitter es muy útil para identificar aquello que las personas quieren saber o encontrar, Twitter ofrece información de lo que sus usuarios están haciendo, sintiendo o pensando.

7.2.2. Selección del sector de actividad

Realizamos un estudio de los sectores de la actividad económica en los que los usuarios se muestran más activos. Tomamos en consideración el sector de las marcas de mayor prestigio dedicadas al gran público consumidor y con fuerte actividad de comunicación publicitaria. Asimismo, planteamos la posibilidad de estudiar los partidos políticos, sector realmente interesante debido a la posibilidad que permite observar posiciones definidas y manifestaciones extremas. Del mismo modo, por ser un terreno controvertido y muy cambiante, preferimos no analizarlo como objeto central de esta tesis.

El sector seleccionado fue finalmente el sector bancario. Una de las razones fue el considerar como apropiada y pertinente la petición del grupo de investigación en estrategia y creatividad en publicidad y relaciones públicas de la Facultad de Comunicación y Relaciones Internacionales Blanquerna de la Universidad Ramón Llull. El interés del grupo de investigación radicaba en la voluntad de conocer las percepciones de los usuarios españoles durante la crisis económica respecto a la banca ya que, precisamente, este sector ha obtenido especial relevancia y presencia en los medios de comunicación y como tema frecuente de debate en la opinión pública.

A esta petición se suma la no menos importante voluntad de demostrar que el SA, a través de una herramienta como SETAL, puede demostrar su utilidad para la elaboración de *insights* publicitarios. Para ello, el mejor sector a analizar debía ser referente al sector empresarial y presentar también una alta necesidad de uso de campañas publicitarias.

Como también comprobamos, la mayoría de los usuarios más activos en las redes sociales mantienen conversaciones sobre una gran cantidad de temas que afectan a sus vidas, tanto a nivel personal como profesional. Todas las conversaciones de los *influencers* pueden ser susceptibles de ser analizadas si se consideran óptimas para la finalidad de nuestro estudio.

7.2.3. Variables técnicas de análisis

VARIABLES DE SEGMENTACIÓN BÁSICAS DE LA INVESTIGACIÓN:

- Palabras clave de selección:
 - Crédito, préstamo, hipoteca, depósito, preferente.
 - Banco, caja de ahorros, banca, financiero
 - La Caixa, BBVA, Santander, Banco Sabadell Atlántico, Catalunya Caixa
 - Se incluyen las variaciones: plurales y errores
- Sector de actividad económica:
 - Bancario y financiero.
 - Destacan comentarios relativos a temas sociales, políticos, económico-financieros y personales.
- Red Social:
 - Twitter
- Tipología de usuarios:
 - *Influencers*: activos regularmente y con 5.000 o más seguidores.
 - Españoles o cuya cuenta de Twitter esté ubicada en España.

Consideramos que con 5.000 cuentas de usuarios la muestra del análisis sería representativa y suficiente.

Con el primer análisis, además de las cuentas de particulares, la tipología de cuentas que encontramos pertenecía al siguiente tipo de entidades:

- Bancos y entidades financieras
- Medios de comunicación: televisiones, periódicos, etcétera
- Partidos políticos
- Otras empresas privadas

Nuestro objetivo en esta tesis es estudiar las conversaciones que mantienen las personas acerca de sus necesidades, opiniones y sentimientos, con lo que decidimos eliminar de la selección las cuentas que no fuesen estrictamente de particulares.

Hecha esta primera primera criba, la lista definitiva de particulares *influencers* fue de 4.543. La lista se confeccionó el 16 de noviembre de 2013.

El proceso de recogida de información se inició el 25 de noviembre de 2013 y el ritmo periódico de recolección y descarga de tuits fue de unos 100.000 tuits semanales hasta que el 30 de marzo de 2014 se alcanzaron los 1.200.000 tuits.

A medida que se iban obteniendo los tuits, se almacenaban en una base de datos para la investigación cuantitativa y que el algoritmo los pudiese identificar.

Los tuits identificados eran contrastados por un diccionario con más de 816 palabras clave relacionadas con la terminología bancaria. De este modo se pudieron identificar los *influencers* que utilizaban dichos términos con mayor frecuencia y también la selección de las conversaciones con mayor nivel de interés. (Ver Anexo: Diccionario del sector bancario).

En esta parte del proceso de análisis, la herramienta clasificó y jerarquizó a los *influencers* de mayor a menor uso de las palabras

A este nivel, la herramienta clasifica y jerarquiza a aquellos *influencers* que utilizan más palabras del diccionario.

Como medida adicional, aplicamos diversos filtros para mejorar la eliminación de “ruido”.

Criterios de selección de los tuits:

- Longitud del tuit superior a 7 palabras.
- Presencia de “#”, “urls”, emoticonos y etcétera inferior al 40% del contenido total de los tuits.
- No consideración de tuits con predominio de palabras irrelevantes para el estudio: saludos, insultos, etc.

Los siguientes filtros que aplicamos fueron de análisis de vocabulario y gramática.

Después de la aplicación de los diferentes filtros, la cantidad final de tuits para el análisis fue de 14.863 tuits, que representa el 1,24% del total. Respecto a los *influencers* que finalmente cumplieron las características suficientes para ser analizados, nos encontramos con un total de 392 (8,63%).

7.2.4. Resumen de los textos de los tuits

Como hemos afirmado, el total de tuits fue finalmente analizados fue de 14.863.

Ordenamos los tuits en base a la puntuación dada por la coocurrencia de las palabras utilizadas, la centralidad y su distribución estadística. Con este orden pudimos ordenar los tuits de mayor a menor importancia y logramos jerarquizar los conceptos y expresiones más relevantes. Junto a la herramienta que determinaba la importancia respecto al contenido y las palabras clave, asociamos a las oraciones una serie de sentimientos.

Además de los tuits, aprovechamos el contenido que el usuario utiliza en su perfil para obtener, partiendo de su “huella gramatical”, la “huella digital”.

7.2.5. Medición de los sentimientos de los tuits

Mediante el diccionario LIWC, mejorado de forma manual por el Dr. Monte para adecuarlo a las expresiones en castellano y catalán, analizamos el cálculo de los sentimientos. Realizada la traducción, las palabras se ordenaron alfabéticamente para identificar las repeticiones y detectar sinónimos que preserven el significado original.

En cuanto a insultos, tacos o palabras cariñosas sin traducción equivalente, buscamos su equivalente en cuanto a significado. El resultado final del diccionario fue de 2.295 palabras asociadas a una carga emocional. (Ver Anexo 8.1.)

1ª Fase:

Obtención de la información:

- Identificación de los influencers = 5.000
- Período de recogida de información = 3 meses
- Cantidad total de tuits = 1.200.000

2ª Fase:

Filtro (fase de segmentación y eliminación del “ruido”)

- Segmentación de influencers (personas físicas) = 4.543
- Selección cualitativa de tuits = 14.863
- Porcentaje de tuits útiles sobre el total = 1,24%
- Tuits que corresponden a influencers = 392
- Influencers analizados sobre el total = 8,63%

3ª Fase:

Análisis gramatical y del sentimiento

- Análisis de los tuits
- Jerarquización
- Polaridad: positiva, negativa y neutra de los tuits
- Puntuación de la variable Sentimiento

La información resultante de la selección, filtro y análisis de los tuits nos ha permitido obtener información relevante como:

- Qué dicen los usuarios
- Cómo se expresan

- Qué palabras clave y qué combinaciones de palabras son las más frecuentes
- Qué emoticonos utilizan con mayor frecuencia
- Las parejas de palabras más frecuentes que utiliza un usuario permite obtener información relevante de las expresiones idiosincrásicas⁷⁶ de los individuos.
- Identificar la polaridad de sentimientos asociados a las horas del día, días de la semana e incluso meses del año. En definitiva, permite saber en qué horas, días y meses, los usuarios son más o menos “felices”.

Población de referencia: 4.543 usuarios analizados.

Contenido total de comentarios: 1.200.000 tuits.

Muestra final: 392 individuos (n= 392).

En las condiciones siguientes, el cálculo del margen de error (error de muestreo) para la distribución en poblaciones finitas (inferiores a 30.000) se realiza con la siguiente fórmula:

$$\sigma_p = \sqrt{\frac{p \cdot q}{n}} \cdot \sqrt{\frac{N - n}{N - 1}}$$

Donde:

- σ_p = Error de muestreo
- p = probabilidad que se produzca el suceso
- q = (1 - p) = probabilidad que no se produzca el suceso
- N = Dimensión de la población
- n = Dimensión de la muestra

Así pues, los márgenes de error para la extrapolación de los resultados sería:

- Para un nivel de confianza del 95,5% (2Q) = +- 4,7% (p=q=50%)

⁷⁶ Idiosincracia: rasgos, temperamento, carácter, etc,.. distintivos y propios de un individuo o de una colectividad. Fuente: RAE. <http://www.rae.es>

7.3. Investigación empírica

7.3.1. Proceso de análisis de los contenidos

7.3.1.1. Criterios de clasificación de las variables de análisis

A continuación pasamos a analizar los resultados obtenidos en la investigación empírica.

Dividimos el fichero resultante en los apartados siguientes:

- Clasificación temporal de los sentimientos:
 - clasificación de las variables: horas del día, semanas y meses en relación a la polaridad de los sentimientos.
- Oraciones más importantes:
 - clasificación de las oraciones por importancia.
- Palabras más importantes (vocabulario):
 - clasifica la totalidad de los textos obtenidos del total de los tuits analizados y ordena las palabras por orden de importancia.
- Segmentación por *topics*:
 - agrupación de palabras clave que aparecen juntas con mayor frecuencia, que tienen significados parecidos y son más centrales en las conversaciones de los usuarios. Esta información aparece en tres dimensiones: palabras clave, selección de oraciones clave que muestran las diferencias entre *topics* y la distribución porcentual de los tuits de los según la pertenencia a los diversos *topics*.
- Usuarios más importantes:
 - clasificación de los usuarios según su importancia.
- Sentimientos asociados a los usuarios:

- clasificación de los usuarios según la puntuación total obtenida en relación a los sentimientos positivos y negativos expresados. El primero de la lista es el usuario cuya suma de sentimientos positivos y negativos es mayor.
- Análisis gramatical de los usuarios:
 - Clasificación según la estructura gramatical de los tuits utilizada por cada usuario. Presentado sin ordenación.

Consideramos relevante recordar la definición de *insight* para agilizar la vinculación entre nuestro proceso de análisis y su vinculación directa con la finalidad de esta tesis: “El término *insight* alberga diferentes conceptos: ideas, motivaciones, vivencias, sentimientos, necesidades y frenos, entre otros. A partir de dichos conceptos se pueden construir nuevos conocimientos que sirvan como desencadenantes de una idea o una acción de comunicación concreta que actúe de guía para el posicionamiento de una marca.”

Con el análisis que estamos presentando y que continuaremos desarrollando en este capítulo podremos presentar una nueva metodología para la elaboración de *insights*.

7.3.1.2. Análisis de las variables de estudio

7.3.1.2.1. Clasificación temporal de los sentimientos

Sentiment en función de la hora del día

Como observamos, los sentimientos expresados en Twitter cambian en función de las horas del día.

Los datos se obtienen a partir de la suma de todos los sentimientos (positivos y negativos) expresados en tramos de una hora. El valor resultante es la resta entre los sentimientos negativos y los positivos.

En el gráfico siguiente podemos observar que los períodos del día donde los usuarios expresan más sentimientos positivos es de 09h a 12h y de 21h a 22h.

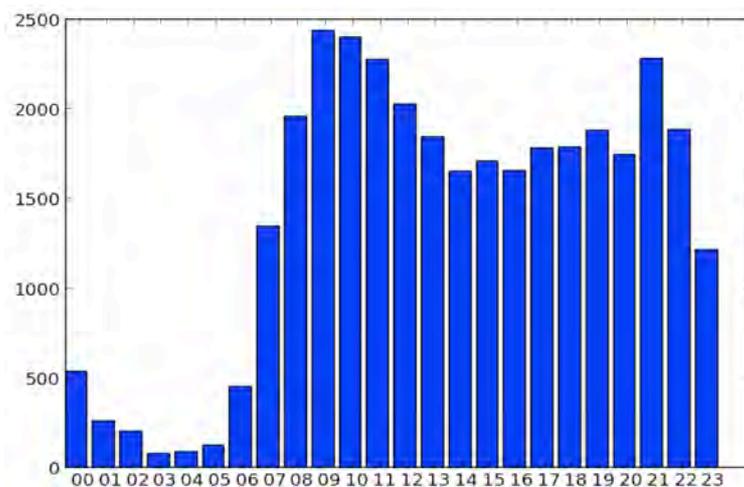


Figura 49: Gráfico de medición del Sentimient en función de la hora del día.

La explicación a la presencia de sentimientos positivos en esas dos franjas horarias responde a que los usuarios tienden a desearse los buenos días y las buenas noches, como es habitual en el uso de hashtags como #BuenosDias, #BuenasNoches o #BuenLunes, #FelizMartes, etc.

Ejemplo de tuit (Código usuario = 114) de estas características, que SETAL puntúa con 6 puntos de Sentimiento positivo y 0 puntos de Sentimiento negativo.

Concretamente este tuit se ha realizado un lunes a las 09:46

@carne2L hola, buenos días, guapisima y feliz lunes;))) | <http://ln.is/blogspot.com/veq2>

Ejemplo de tuit (Código usuario = 1) de estas características, que se puntúa con 10 puntos de sentimiento positivo y 0 puntos de sentimiento negativo. Concretamente, este tuit se ha realizado un viernes a las 06:43.

“@lucy_arties Buenos días Lucí, que tengas un viernes bonito, sonríte y se feliz, un beso.”

Ejemplo de tuit (Código usuario = 140) de estas características, que se puntúa con 8 puntos de Sentimiento positivo y -2 puntos de Sentimiento negativo. Concretamente, este tuit se ha realizado un lunes a las 21:07.

“Siempre hay una segunda oportunidad para hacer lo q realmente queremos lograr en nuestro paso por el mundo. No pierdas la fe. Feliz #martes”

Sentiment en función del día de la semana

En vez de realizar la agrupación por horas, en esta ocasión se hizo según cada día de la semana. Observamos que el lunes y el martes son los dos días con menos expresión de sentimientos, siendo el martes el menor de la semana. Observamos que el domingo es bastante neutro y muestra una caída de expresión de sentimientos, como si acusase también el inicio de la semana, el regreso al trabajo y el final del período de descanso. Observaremos que, a medida que se acerca el fin de semana, la expresión de sentimientos es mayor, llegando al viernes como día álgido.

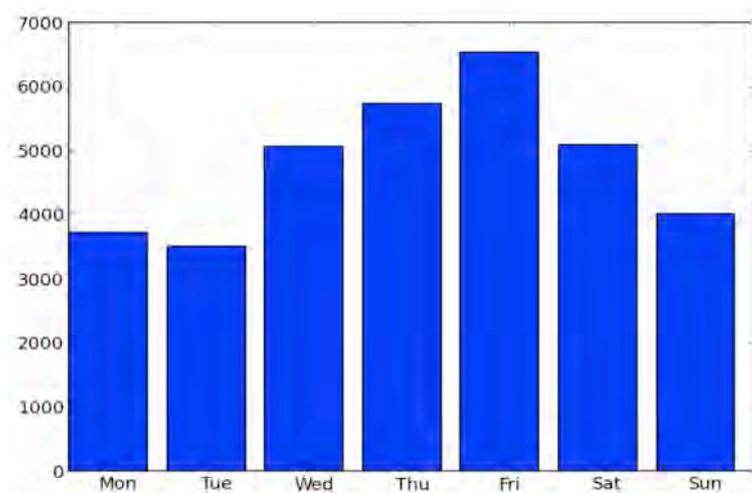


Figura 50: Gráfico de medición del Sentiment en función del día de la semana.

Ejemplo de tuit (Código usuario = 176) de estas características, puntuado con 10 puntos de sentimiento positivo y 0 puntos de sentimiento negativo.

Concretamente, este tuit se ha realizado un viernes a las 09:49.

“@CandidoConde buenos días, seguro q será muy interesante, lo veremos, y mi apoyo para vosotros, abrazo”

Ejemplo de tuit (Código usuario = 1) de estas características, puntuado con 10 puntos de sentimiento positivo y 0 puntos de sentimiento negativo.

Concretamente, este tuit se ha realizado un viernes a las 06:43.

“@lucy_arties Buenos días Lucí, que tengas un viernes bonito, sonríte y se feliz, un beso.”

Ejemplo de tuit (Código usuario = 21) de estas características, que se puntúa con 11 puntos de sentimiento positivo y 0 puntos de sentimiento negativo.

Concretamente, este tuit se ha realizado un viernes a las 17:52.

“Acaba un año maravilloso tanto a nivel personal como laboral. Espero que el vuestro haya sido aún mejor. Feliz año por anticipado”

Sentiment en función del mes del año

Realizamos un proceso análogo al llevado a cabo con los días y los meses. En general, observamos que existe una diferencia respecto a la expresión de sentimientos en función de los meses, así como una tendencia decreciente de sentimientos positivos a partir de diciembre y con el avance del invierno.

En nuestro trabajo, como señalamos en la determinación de la muestra, estudiamos los meses comprendidos entre noviembre y marzo, es decir, diciembre, enero y febrero de 2014 totalmente completos.

De los tres meses analizados, determinamos diferentes niveles de expresión de sentimientos por parte de los usuarios, siendo diciembre el mes más expresivo y febrero, el menor. Lo ilustramos con el gráfico siguiente.

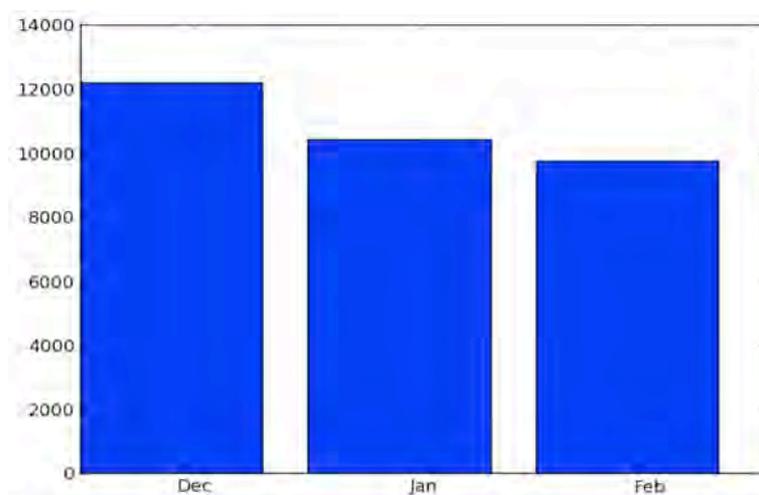


Figura 51: Gráfico de medición del Sentiment en función del mes del año

Con el estudio de estos datos estacionales, empezando por la distribución por horas y alcanzando la distribución mensual de la expresión de los sentimientos, concluimos que la repercusión en campos de la comunicación (como el marketing) puede resultar muy elevada. Con esta información se pueden llegar a determinar qué momentos (horas, días o meses) son mejores para según que comunicaciones, ofertas o promociones, entre otras acciones. Sin duda, la recepción de los usuarios será diferente en cada momento del día y, por ende, los resultados obtenidos con las campañas publicitarias también presentarán variaciones considerables. Por ejemplo, es conocida la relación positiva entre un estado de ánimo optimista y la predisposición a la compra.

7.3.1.2.2. Oraciones más importantes

Mostramos a continuación la clasificación de las oraciones en función de su relevancia. Recordamos que estudiamos finalmente 14.863 de un total de 1.200.000.

Expresamos la importancia de cada oración de forma porcentual. La más importante obtiene tan solamente un 0,03119% de importancia y la menos importante obtiene un 0,00000%.

Debido a la magnitud del fichero expresado en porcentajes, la importancia de cada oración parece mínima y poco significativa, pero la observación es radicalmente diferente si las comparamos entre sí veremos como la más importante es 135,60 veces más importante que la última (puntuación de la última oración = 0,00023%).

El ahorro económico y en coste temporal de poder realizar esta clasificación de forma automática es considerable y merece ser tenido en cuenta. Confirmamos aquí los puntos señalados en capítulos anteriores acerca de las ventajas de la aplicación de métodos RAT respecto a los sistemas tradicionales.

A continuación, presentamos la información obtenida a través de una selección de los tuits más importantes según los criterios establecidos.

Tuit del usuario número 83, a partir de la lectura de un artículo del sitio web relacionado con temas de la abogacía:

“La ejecución hipotecaria bajo sospecha I ¿Hay Derecho?”

Importancia en % = 0,03119%. Sentimiento positivo= 0. Sentimiento negativo= 0

Este tuit sube en importancia debido a la coincidencia entre las palabras “hay derecho” y la web del mismo nombre. Aparece muy a menudo y de forma central en muchos tuits.

Tuit del usuario número 203, a partir de la lectura de un artículo de un blog de un sitio web relacionado con temas de la abogacía:

“No hay buen abogado sin vocación y sin pasión por su oficio pese a las muchas adversidades”

Importancia en % = 0,02305%. Sentimiento positivo= 0. Sentimiento negativo= -2

Tuit del usuario número 6, a partir de la lectura de un artículo de un blog de un sitio web relacionado con temas de economía:

“Caja España: “Hay que vender preferentes, como sea. Nos va la vida en ello” y hasta hoy....”

Importancia en % = 0,02047%. Sentimiento positivo= 2. Sentimiento negativo= 0

Demuestra y comparte su malestar por lo que considera una práctica profesional poco ética.

Tuit del usuario número 150, a partir de la lectura de un artículo divulgado y compartido a través del sitio web meneame.net.

“No es ya: vaya cagadas que hace el gobierno del PP!! Es vaya cagada del gobierno del PP! Iros ya coño!”

Importancia en % = 0,02017%. Sentimiento positivo= 0. Sentimiento negativo= -2

Comentario con carga emocional claramente negativa en relación a la gestión realizada por el gobierno.

Tuit del usuario número 322, a partir de la lectura de un artículo de la sección de política de la versión digital del diario El País:

“Para esto si hay dinero. CIU se queda sola en la defensa del simposio “España contra Cataluña”

Importancia en % = 0,01986%. Sentimiento positivo= 1. Sentimiento negativo= -3

Tuit del usuario número 174:

“La #república no va a venir. Dejen de soñar. Hay que ir a por ella”

Importancia en % = 0,01954%. Sentimiento positivo= 0. Sentimiento negativo= -2

Tuit del usuario número 220:

“¿No hay dinero? 49 millones de euros en helicópteros de Defensa comprados el último día del año”

Importancia en % = 0,01839%. Sentimiento positivo= 0. Sentimiento negativo= -2

Muestra y comparte su disconformidad con la adquisición.

Tuit del usuario número 37:

“¿Hay que dar gracias a Rajoy por las 4 subidas consecutivas de la luz? Según él hay que dárselas porque en Enero subirá menos de un 3%”

Importancia en % = 0,01837%. Sentimiento positivo= 2. Sentimiento negativo= 0

Muestra y comparte su disconformidad con las subidas de la luz.

Tuit del usuario número 225, en relación a los disturbios ocurridos en Burgos en enero de 2014:

“Hay fuego en las calles. Hay lucha en nuestros corazones. #ArdeBurgos”

Importancia en % = 0,01806%. Sentimiento positivo= 0. Sentimiento negativo= -4

Tuit con alto contenido emocional negativo: -4.

Tuit del usuario número 10:

“Hay días en que no puedo más con la mierda eclesiástica y franquista que nos invade. Hoy es uno de ellos”

Importancia en % = 0,01804%. Sentimiento positivo= 0. Sentimiento negativo= -2

Tuit con alta carga emotiva y negativa.

Si nos centrásemos en una empresa en concreto y en las comunicaciones que realizan sus clientes, la información, ordenada jerárquicamente, puede resultar fundamental para medir si los clientes se muestran positivos o negativos respecto a sus productos, gestión o últimas acciones realizadas.

Asimismo, la jerarquización ofrece qué palabras clave y qué aspectos de la comunicación son los más importantes. Ello puede ofrecer una información muy valiosa para poder establecer un diálogo con los clientes.

La satisfacción, el malestar o el interés demostrado por los usuarios que hemos mencionado (como ejemplo de entre todos los tuits analizados) nos otorga, en nuestro propósito destinado al hallazgo de *insights*, un claro espectro de sensibilidades y temas de atención que podríamos utilizar en nuestro favor.

7.3.1.2.3. Palabras más importantes

Mostramos la clasificación del vocabulario expresado por orden de importancia. SETAL, siguiendo el criterio antes detallado, ordena las palabras por orden de importancia y no por mayor número de repeticiones.

En el cuadro siguiente aparecen las palabras por su raíz, su porcentaje correspondiente y las palabras que pueden asociarse a dicha raíz.

Raiz Pal	Imp. %	Pal. asociadas
ris	2,4971%	risas risa
via	1,9732%	vía
espan	0,9976%	espana
anos	0,9405%	anos
millon	0,8416%	millones millón millon
gobiern	0,8272%	gobierno gobiernan gobiernos gobierne gobierna
ano	0,7841%	ano
nuev	0,7755%	nuevas nuevamente nueves nuevos nuevo nueve nueva
cas	0,7067%	casa casado cas caso casada casé casarse casadas casas cases caser casos
buen	0,6580%	bueno buenas buena buenos buenas buen
hoy	0,6552%	hoy
pas	0,6538%	pase pasarse pasa pasadas paso pasé pasarlo pasó pasos pasarle pasamos pasen pasan pasado paseaban pasaban pasará pasas pasarían pases pasar pasaremos paseando pasando pasáis pasaba pasada pas pasarme pasaría pasaron pasara pasándolo pasaréis pasándolas pasados pasarán pasaran
mejor	0,6107%	mejores mejorables mejorando mejore mejoran mejorara mejoró mejorada mejoren mejoro mejor mejoría mejoras mejorar mejora mejorarlo
ley	0,5883%	ley leyen leyes
mes	0,5776%	mesas mesías mes meses mesa más mese

pp	0,5660%	pp
quier	0,5590%	quiero quieren quieran quiere quiera quieras quieres
dic	0,5533%	dicen dic diciendo dice dices
sorpres	0,5045%	sorpresas sorpresivamente sorpresa
banc	0,4719%	bancaría banca banco bancos
pag	0,4691%	pagaría pagarlo pagarle pagarla pagarse pagamos pague pagarán pagando pagues paguen pagaban pagarles pagáis pagar pagas pagarme pagaron pagase pagado pagad pagaba pagan pagada pagarnos pagaré pagarlas paguemos pagará pagó pagos pagara pagarían pagáramos pagados pago pagaremos paga page
gran	0,4611%	granado grano granada gran granados
feliz	0,4584%	feliz
madridcapital	0,4561%	madridcapital
simbolomoned	0,4491%	simbolomoneda

Tabla 5: Tabla de presentación de las palabras más importantes (25 primeras)

La primera de las palabras, más presente y central, es la palabra de raíz “ris”, que corresponde a la traducción automática de la interpretación de los emoticonos. En este caso, corresponde al emoticono asociado a la risa o risas, con un 2,50%.

7.3.1.2.4. Segmentación por *topics*

Los *topics* nos permiten detectar qué temas preocupan más a los usuarios o por los que manifiestan un mayor interés.

Identificamos cuatro grupos de *topics* con sus palabras clave asociadas y las oraciones más destacadas según nuestros criterios.

Presentamos a continuación una relación de los *topics*, oraciones clave y distribuciones porcentuales de cada usuario.

NomUsuari	TOPIC 1	TOPIC 2	TOPIC 3	TOPIC 4
Pal Clave	derechoes ía vía espana anos quieres gobierna nueva vidas hoy pp casos dices	buen feliz sorpresa dia hoy díasen gracia mejorarlo semana veremos pasaran casos amiga saludaban	espana anos gobierna millon ano nueva pasaran dices mese pp leyes quieres deb luz casos	elpais eldiari casos millon gobierna públicos elmundoes madridcapital anos nueva mese

Tabla 6: Tabla de *topics* por palabra clave

Topic 1

Representa textos sobre temas políticos, relacionados con el partido en el gobierno (PP), derechos, España y en cómo afecta a la cotidianidad de los españoles.

Presentamos en la figura diversas oraciones con el *topic 1*.

Ora Clave	RT @AsocDRYEventos: Sólo seis banqueros se enfrentan al banquillo de los acusados en cinco años de crisis #TodosEnLa... http://ow.ly/tycEp
Ora Clave	@DailyMur @diariolaopinion Y todo por decir a Rajoy al rostro que se tirara al río; esto es, que se fuera a la shit que aquí no le queremos http://bit.ly/1jH3Dnc
Ora Clave	Firma y divulga. ELLOS SON LOS QUE NOS DEBEN DINERO... http://t.co/C0ovm3zCTD http://fb.me/38JuHL3Xr
Ora Clave	¿cuántos años de salario tienes que destinar para comprar una casa según la provincia en la que vivas? http://t.co/6VwvYewePh http://bit.ly/1eJNY4P

Tabla 7: Tabla de tuits del topic 1

Topic 2

Agrupamos aquellos contenidos de tuits con predominancia de la relación entre amigos y familiares. Contrariamente al *topic 1*, no son reivindicativos ni agresivos en sus contenidos. De hecho, menudo son tuits de agradecimiento, saludo, diversión o que comparten alguna información útil y sin voluntad de crítica ni de denuncia. En otras ocasiones se limitan a compartir o retuitear un artículo de ésta índole.

Ora Clave	¿Dejarías de trabajar si te toca un buen premio en la lotería del Gordo de #Navidad? http://t.co/dNS2nUBT68 http://bit.ly/1fmwDym
Ora Clave	Muchas gracias a la gran familia de @PAH_Madrid por hacernos sentir importantes, pero cuidao que se nos sube a la cabeza! Muchos abrazos! http://www.cadenaser.com/espana/articulo/ignacio-gonzalez-expresa-sorpresa-estupor-imputacion-mujer/csrrpor/20131218csrrsmac_10/Tes
Ora Clave	RT @EstrategiasTv: #FlujosDeFondos Los fondos vuelven a la renta fija tras cinco semanas. http://t.co/8dSUds2zmZ http://ow.ly/ssQTr
Ora Clave	RT @SurFincas: Última oportunidad para pagar menos en la Declaración de

	la Renta http://t.co/0sz9bnUJvT via @Vivienda_ELPais http://bit.ly/1dIMkif
--	---

Tabla 8: Tabla de tuits del topic 2

Topic 3

Este topic agrupa contenidos de tuits que pueden compartir, en algunos casos, coincidencia con los contenidos y las palabras identificadas en el *topic 1*. Sin embargo, la mayor diferencia que presenta respecto al *topic 1* es el tono, más suave y menos agresivo que en este *topic 3*.

Ora Clave	RT @ElviraMeligrana: Lo de la sede del #PP no es un registro, es que son tan buenos anfitriones que la policía no ve la manera de irse. Cor... http://youtu.be/8vTGDROAGrA
Ora Clave	RT @DolorsAlvarez: “me vais a curar el cancer, pero me matareis de un infarto esperando seis horas la ambulancia para volver a casa” #oidoe... http://goo.gl/UIClil
Ora Clave	RT @TaboadaLucia: En las búsquedas de Google de la infanta Cristina ya sale “voy a tener suerte” y “estoy convencida de que me va a ir bien... http://www.elmundo.es/internacional/2014/01/23/52e10167e2704ecd6c8b456d.html?a=72c814f9d2cb6325ac6c15832d784630&t=1390480536
Ora Clave	RT @danicamposd: URGENTE: El Constitucional avala la reforma laboral del Gobierno de Rajoy por siete votos a favor y cuatro en contra https://www.google-sciencefair.com/es/
Ora Clave	RT @ComunicAusbanc: Condenado un banco a devolver 108.000 euros gracias a @Ausbanc http://t.co/BKwcoPBFUq vía @CORDOBA_diario http://www.diariocordoba.com/noticias/cordobalocal/condenado-banco-devolver-108-000-euros_862505.html

Tabla 9: Tabla de tuits del topic 3

Topic 4

Selección de contenidos de los tuits que agrupan palabras clave relacionadas con la banca, el gobierno, la política y otros temas de interés general. Normalmente, se limitan a reproducir los titulares de los artículos aparecidos en periódicos digitales (el pais.com, que.es, El Mundo, etcétera). Hay una gran masa de retuits y observamos cómo los usuarios siguen hablando del gobierno pero a partir de los artículos que consumen.

Ora Clave	RT @AciertoCom: Sigue el culebrón de la luz. El Tribunal Supremo decreta que las eléctricas deben financiar el déficit en solitario http://... http://elpais.com/elpais/2013/12/11/gente/1386787623_500516.html
Ora Clave	El gobierno aprueba el nuevo índice hipotecario que entrará en funcionamiento el 1 de noviembre http://t.co/c5TphMFysS http://ow.ly/tOgi1
Ora Clave	Lo de “el coñazo del desfile” del carallo son palabras del gran patriota Mari Rajoy a su protector PP María Asnar. http://t.co/AhwT0s4dOP http://www.que.es/ultimas-noticias/espana/201210120900-hispanidad-abucheos-zapatero-conazo-rajoy-cont.html
Ora Clave	Una buena noticia para las hipotecas El euribor podría bajar al 0,30%, con un ahorro medio anual de 156 euros para las hipotecas http://ow.ly/tOgi1
Ora Clave	Pobre Gallardón. Ahora va el socio de Urdangarín y pide que declare en el juicio. Sí hombre, ni que fuera un civil normal ante la ley! http://www.despacitoybuenalettra.com/

Tabla 10: Tabla de tuits del topic 4

Tanto las fuentes de información que tienen los usuarios como los artículos que compartimos aportan una valiosa información acerca de su forma de ser, inclinaciones y temas de interés. Gracias al SA que estamos realizando, la importancia es todavía mayor al poder asociar los sentimientos a la información que afirman consumir y sobre la que opinan públicamente, ya sea mostrando su disconformidad o su empatía.

7.3.1.2.5. Distribución porcentual de los tuits por *topics*

Si analizamos a los usuarios de Twitter a partir de los *topics*, aparecen sus conversaciones distribuidas porcentualmente del modo siguiente:

La columna de la izquierda corresponde al código de cada usuario y los datos (expresados en porcentajes) representan la distribución de los textos según diversas palabras agrupadas por *topics*.

Código de usuario	TOPIC 1	TOPIC 2	TOPIC 3	TOPIC 4
	derechoes ía vía espana anos quieres gobierna nueva vidas hoy pp casos dices	buen feliz sorpresa dia hoy díasen gracia mejorarlo semana veremos pasaran casos amiga saludaban	espana anos gobierna millon ano nueva pasaran dices mese pp leyes quieres deb luz casos	elpais eldiari casos millon gobierna públicos elmundoes madridcapital anos nueva mese
0	23,17%	22,24%	43,81%	10,78%
1	9,49%	84,38%	4,88%	1,25%
2	25,70%	11,82%	59,15%	3,34%
3	13,22%	17,52%	33,29%	35,97%
4	18,89%	11,05%	18,15%	51,90%
5	7,47%	5,13%	83,87%	3,53%
6	25,42%	35,38%	32,33%	6,87%
7	1,07%	3,98%	61,47%	33,48%
8	11,22%	19,78%	63,35%	5,65%
9	32,42%	16,50%	39,46%	11,62%
10	28,98%	13,43%	34,46%	23,13%
11	5,11%	9,26%	52,91%	32,72%
12	25,69%	39,05%	33,57%	1,69%

13	30,24%	18,52%	47,73%	3,51%
14	14,23%	8,68%	71,33%	5,75%
15	12,09%	59,43%	23,06%	5,42%

Tabla 11: Tabla de distribución de los tuits por topics

En el cuadro anterior hemos seleccionado los 16 primeros usuarios, correspondientes a los códigos del 0 al 15. La agrupación representa los contenidos de todos los tuits realizados por dichos usuarios durante el período del estudio. La suma de todos los *topics* es igual al 100% del total de tuits.

7.3.1.2.6. Ejemplos de tuits según sus *topics*

Topic 1

El usuario con código 41 concentra el 49,54% de sus tuits en el *topic 1*.

Habla sobre temas políticos y bancarios desde una óptica más reivindicativa:

Tuit en relación a una noticia que implica al Ministerio de Justicia

“estas declaraciones son mas desternillantes que las del PP diciendo q estamos saliendo de la crisis gracias al esfuerzo de los q mas tienen”

Tuit en relación a la crisis bancaria:

“todavía la gente sigue guardando (quien pueda llegar a tener) su dinero en un banco/caja/sus muertos?”

Tuit que refleja su personalidad:

“ah y otra cosa, a los q no creyeron...yo soy o blanco o negro,, gris no voy a ser en mi vida”

Topic 2

El usuario con código 1 concentra el 84,38% de sus tuits en el *topic 2* :

Efectivamente, utiliza Twitter para relacionarse con sus amistades (dando los buenos días, festejando celebraciones, organizando el fin de semana con los amigos, entre otros)

A modo ilustrativo, señalamos este tuit:

“@Xxx_xxxx Buenos días Lucí que tengas un bonito domingo y mires al futuro con alegría, un beso amiga.”

Topic 3

El usuario con código 68 concentra el 66,42% de sus tuits en el *topic 3*.

Recordamos que, si bien muchos tuits comparten temática y objeto con el *topic 1* (gobierno, leyes, España, PP, año, etcétera), su tono es sustancialmente diferente, sin el uso tan contundente de expresiones reivindicativas ni agresivas.

“El Gobierno quiere que la banca asuma pérdidas en el rescate de las autopistas”. Eso significa, más rescates a la banca. Sistema absurdo”

Retuit de un artículo de el diario Expansión:

“El Banco de España y Bruselas, los que menos aciertan en la diana de las previsiones económicas”

Retuit de un artículo de la sección de economía de El País, con comentario añadido:

“La sobrina de Guindos irá al Banco Mundial y ganará 13.000 euros al mes <http://t.co/7QbA02jevi> Pero no es por enchufe, ehhhh”

Topic 4

Encontramos a usuarios que mayoritariamente retuitean artículos de prensa sobre temas concretos: vivienda, banca o política. En ocasiones comentan las noticias que retuitean.

El usuario con código 60 concentra el 71,67% de sus tuits en el *topic 4*.

Se reconoce por ser un usuario asiduo a la lectura de periódicos (“eldiario.es”, “publico.es”, “laverdad.es”, “elpais.es”, entre otros).

No realiza retuits de otros usuarios, pero sí reproduce los titulares de las noticias que considera más relevantes de distintos ámbitos de la política y sociedad actuales, como por ejemplo:

“Ricardo Sixto: “Hay que investigar las expulsiones ‘en caliente’ de este y del anterior Gobierno” <http://t.co/O8QHP2x9rB> vía @eldiarioes | http://www.eldiario.es/_dbf1a19”

“El Gobierno regala hoy a la banca hasta 30.000 millones de euros en créditos fiscales <http://t.co/93AoqMzz6X> vía @publico_es | <http://www.publico.es/dinero/485670/el-gobierno-regala-hoy-a-la-banca-hasta-30-000-millones-de-euros-en-creditos-fiscales>”

“Bernal diseña un ajuste de 300 millones para su futuro sustituto. La Verdad <http://t.co/afyKPhGzfw> vía @laverdad_es | <http://www.laverdad.es/murcia/v/20140220/region/bernal-disena-ajuste-millones-20140220.html>”

“Los consejeros de Caja Madrid tenían tarjetas de crédito en ‘negro’ de hasta 50.000 euros al año <http://t.co/z9dbQAuwUn> vía @eldiarioes | http://www.eldiario.es/_c53fa2b”

De este modo, el usuario comparte con sus seguidores las noticias que quiere contribuir a aumentar su difusión y, al mismo tiempo, le permiten transmitir su opinión a través de los artículos de los periodistas.

7.3.1.2.7. Análisis de los usuarios por *topics*

Profundizando más en la información que nos ofrece la clasificación y ordenación de los usuarios por *topics* procedemos a analizar a aquellos usuarios (10 primeros) que concentran sus tuits en cada uno de los cuatro *topics*.

Pensamos que, agrupados como segmento, pueden aportar una información adicional respecto a presentar la información separa por individuos.

Topic 1

Selección por orden descendiente de los diez primeros usuarios cuyos tuits se concentran en el *topic 1*:

Cod	Topic 1	Topic 2	Topic 3	Topic 4
374	70,05%	7,06%	15,43%	7,46%
377	64,29%	11,02%	22,01%	2,69%
337	61,04%	10,14%	26,40%	2,42%
351	56,82%	29,91%	12,03%	1,23%
390	56,74%	27,53%	12,30%	3,43%
296	54,17%	31,22%	14,18%	0,43%
283	53,84%	7,33%	21,79%	17,05%
262	53,73%	7,47%	26,18%	12,63%
346	53,61%	20,13%	24,80%	1,47%
301	53,12%	9,38%	34,56%	2,94%
Promedio	57,74%	16,12%	20,97%	5,17%

Tabla 12: Tabla de ordenación de los usuarios por *topic 1*

Observamos que los usuarios cuyo *topic* mayoritario es el 1 tienen mayor relación con el *topic 3* y menor relación con los *topic 4* y 2. Según el análisis realizado de las agrupaciones

de palabras y contenidos de los tuits, esta relación es lógica: los usuarios que utilizan Twitter para sus reivindicaciones también realizan algunos tuits que son de la misma temática pero que no son tan agresivos.

Es lógico observar que los usuarios contenidos en el topic 1, debido a su fuerte generación de contenidos propios, sean los que menos contenido externo retuitean y, por lo tanto, tengan poca relación con el *topic 4* (más asociado a retuits).

Topic 2

Selección por orden descendiente de los diez primeros usuarios cuyos tuits se concentran en el *topic 2*.

Cod	Topic 1	Topic 2	Topic 3	Topic 4
59	16,54%	77,92%	5,16%	0,38%
311	7,91%	73,79%	14,75%	3,55%
354	3,52%	72,43%	21,89%	2,17%
297	17,41%	59,27%	21,04%	2,28%
76	18,63%	55,33%	22,14%	3,90%
382	3,16%	55,19%	39,17%	2,48%
358	2,25%	53,76%	32,90%	11,09%
349	3,62%	53,51%	41,53%	1,34%
114	25,54%	53,12%	13,09%	8,25%
240	1,35%	52,43%	38,08%	8,13%
Promedio	9,99%	60,68%	24,97%	4,36%

Tabla 13: Tabla de ordenación de los usuarios por topic 2

A partir de las características analizadas anteriormente de los usuarios del *topic 2* constatamos la lógica al observar su vinculación al topic 3: siguen hablando de temas de interés nacional y general, pero con un tono menos agresivo que los usuarios del *topic 1*, con el que no se sienten identificados para expresarse.

Del mismo modo que los usuarios del *topic 1*, coinciden poco con el *topic 4* porque sus contenidos expresan sus propias emociones y sentimientos y no dependen de los retuits de medios de comunicación digitales.

Topic 3

Selección por orden descendiente de los diez primeros usuarios cuyos tuits se concentran en el *topic 3*.

Cod	Topic 1	Topic 2	Topic 3	Topic 4
241	1,77%	10,05%	86,73%	1,45%
20	6,68%	7,57%	80,14%	5,62%
86	9,88%	8,61%	80,07%	1,43%
178	14,33%	3,52%	79,20%	2,95%
110	11,53%	10,04%	76,77%	1,67%
238	13,51%	8,47%	75,52%	2,49%
99	17,06%	7,33%	74,28%	1,32%
385	17,47%	8,33%	73,44%	0,76%
19	1,56%	21,48%	73,14%	3,82%
38	17,43%	6,66%	73,13%	2,78%
Promedio	11,12%	9,21%	77,24%	2,43%

Tabla 14: Tabla de ordenación de los usuarios según el *topic 3*

De igual modo que los usuarios del *topic 1* coincidían con los del *topic 3*, se produce la misma relación pero en sentido contrario. Comparten relaciones de interés por los mismos contenidos y puntualmente pueden utilizar un tono más agresivo. La menor relación la mantienen los usuarios ubicados en el *topic 4*.

Topic 4

Selección por orden descendiente de los diez primeros usuarios cuyos tuits se concentran en el *topic 4*.

Cod	Topic 1	Topic 2	Topic 3	Topic 4
195	6%	1%	8%	85%
341	1%	7%	12%	80%
125	4%	2%	22%	72%
60	7%	3%	18%	72%
273	1%	5%	26%	68%
187	16%	1%	18%	65%
214	14%	8%	13%	65%
98	11%	5%	22%	62%
17	13%	11%	15%	61%
174	13%	7%	19%	60%
Promedio	8,75%	5,09%	17,23%	68,94%

Tabla 15: Tabla de ordenación de los usuarios según el topic 4

La relación del *topic 4* respecto a los otros se encuentra más sólida si la vinculamos con el *topic 3*. Definimos que la mayoría de sus tuits proceden de titulares de noticias de periódicos digitales, los cuales reflejan de forma periódica y habitualmente neutra los temas de mayor interés. Estos temas compartidos por los usuarios del *topic 4* coinciden con los del *topic 3*, a diferencia de que los usuarios del *topic 3* suelen añadir sus propios comentarios en lugar de simplemente retuitearlos como prefieren los usuarios contenidos en el *topic 4*.

7.3.1.2.8. Usuarios más importantes

Clasificamos a los usuarios según su importancia. La suma de la variable “importancia” del total de los 392 usuarios analizados es igual al 100,00%. El usuario más importante tiene un 0,768% de importancia y el menos importante un 0,019%. La magnitud del fichero hace que los porcentajes que reflejan la importancia de cada usuario parezca diluida, pero si analizamos la diferencia entre el primero y el último observaremos que el primero es 40,42 veces más importante que el último.

Cod usu	Imp%	Senti		Fecha	Oración Central					
		+	-							
83	0,768%	29	-46	Wed Dec11 23:08:14	La ejecución hipotecaria bajo sospecha. ¿Hay Derecho?. http://t.co/C3q516cWKN vía @HayDerecho					
Palabras Clave			Tot Adj dif	Tot Adv dif	Tot Conj dif	Tot Conj Sub	Tot Verbs	Tot Oracio	Long prom oracio	Tot Pal
Risa derecho			33	8	4	3	41	53	19,06	1010

Tabla 16: Tabla del usuario más importante (Ver Anexo 8.3.)

En este gráfico se puede observar la información obtenida del usuario con código 83 (más importante) cuyos títulos de las columnas significan lo siguiente:

- Cod Usu. = Código de usuario
- Imp % = importancia del usuario
- SentiPos = puntuación obtenida de sentimientos positivos
- SentiNeg = puntuación obtenida de sentimientos negativos
- Data = hora, minutos, segundos, día, mes y año en que se realizó el tuit.
- Oració central = Tuit central entre todos los tuits realizados por este usuario
- Paraules clau = palabras clave más centrales e importantes entre todas las utilizadas por este usuario.

- Tot Adj dif = Cantidad de adjetivos diferentes utilizados por este usuario
- Tot Adv dif = Cantidad de adverbios diferentes utilizados por este usuario
- Tot Conj dif = cantidad de conjunciones diferentes utilizadas por este usuario
- Tot Conj Sub. = Cantidad de conjunciones subordinadas utilizadas por este usuario
- Tot. Verbs = cantidad de verbos diferentes utilizados por este usuario
- Total Oraci = cantidad de oraciones totales utilizadas
- Long.pron.oració. = Longitud promedios de las oraciones escritas o retuiteadas por este usuario.
- Tot pal. = Cantidad total de palabras utilizadas en el total de tuits analizados de este usuario.

Análisis del usuario más importante

SETAL ha determinado que el usuario 83 es el más importante sobre temas relacionados con la banca por el siguiente tipo de contenidos:

“La ejecución hipotecaria bajo sospecha | ¿Hay Derecho?”

“Caja Segovia: el increíble caso de la caja menguante (II): | ¿Hay Derecho?”

“¿Qué eran exactamente los “Valores Santander”? | ¿Hay Derecho?”

“El Banco de España habla de mejorar sus procedimientos de supervisión”

“La expulsión de las cajas de ahorros de su paraíso financiero | ¿Hay Derecho?”

“Banco Central de los bancos centrales advierte de un crash en el sistema financiero”

“La inversión en capital riesgo se desploma a la mitad en dos años “

En los tuits mostrados como ejemplo, encontramos una importante presencia de palabras clave relacionadas con temas bancarios: “hipotecaria”, “Caja”, “Valores Santander”, “Banco de España”, “cajas de ahorros”, “financiero”, “Banco Central”, “inversión”, “capital riesgo”.

Estas palabras no son solamente palabras repetidas sino que son centrales en su discurso y además están vinculadas con otras palabras también consideradas “importantes”, lo que hace a este usuario especialmente relevante para el segmento de usuarios analizado.

Comprobamos, en consecuencia, que el criterio de ordenación jerarquizando a los usuarios por su importancia es correcto.

7.3.1.2.9. Sentimientos asociados a los usuarios

Desplegamos la clasificación de los sentimientos positivos y negativos expresados por parte de los usuarios en función del grado de su emotividad. Mostramos a continuación un cuadro con los usuarios jerarquizados en función del valor de los sentimientos expresados.

Num	Senti+/Sent-	Cod Usuario	Sent+	Cod Usuario	Sent-	Cod Usuario
1	321	12	186	12	-167	24
2	309	24	171	76	-150	33
3	267	6	152	6	-147	9
4	266	33	142	24	-146	124
5	263	9	124	59	-142	22
6	240	25	123	1	-140	25
7	237	13	121	13	-136	17
8	219	31	119	15	-135	12
9	218	22	119	114	-133	39
10	217	88	116	9	-128	31
11	211	23	116	33	-126	30
12	211	97	115	106	-122	73
13	207	17	112	0	-119	48
14	200	129	112	88	-116	13
15	195	73	112	140	-115	6

16	193	41	106	102	-115	28
17	193	120	105	97	-115	108
18	192	106	101	27	-114	120
19	191	30	100	23	-112	61
20	190	66	100	25	-112	95
21	188	34	97	47	-111	23
22	188	61	97	129	-110	51
23	188	76	96	34	-106	41
24	188	124	95	194	-106	66
25	187	95	94	45	-106	97

Tabla 17: Tabla que ordena a los usuarios con mayor carga emocional

Observamos que el usuario número 12 suma 321 puntos (procedentes de la suma de los 186 puntos de sentimientos positivos y 135 puntos de sentimientos negativos). Su expresión de sentimientos positivos supera a los negativos en un 37,8%.

Analizamos algunos tuits representativos del usuario 12:

*“Hay cosas k parecen chistes pero en España son el pan d cada dia:
<http://t.co/mnQYHyZBe7> “*

Importancia = 0,01800%. Sentimiento positivo = 2. Sentimiento negativo = 0.

*“@oscarxxxx k mejor motivación k ayudar al peque a mejorar su rendimiento escolar
mientras pasas buenos momentos con él “*

Importancia = 0,01562%. Sentimiento positivo = 8. Sentimiento negativo = 0.

“@Pablojxxxx si tiene ese ratio muy bueno, y muchos otros buenos “

Importancia = 0,00739%. Sentimiento positivo = 7. Sentimiento negativo = 0.

“@Herrxxx bueno es otra manera d aumentar liquidez sin tener k cerrar posiciones, eso si, si sale bién claro”

Importancia = 0,00473%. Sentimiento positivo = 6. Sentimiento negativo = 0.

Por otro lado, si analizamos algunos tuits con mayor expresión de sentimientos negativos, podemos ver:

“@Herrxxx Ence como comenté estos días pasados pinta mal, la regulación afectara mucho. Duro sino salen noticias no tira, raro.”

Importancia = 0,00622%. Sentimiento positivo = 2. Sentimiento negativo = -6.

“Harto d ver tanto consenso y tantos analistas preveiendo subidas d dos dígitos en renta variable. Mala señal, sin duda”

Importancia = 0,00469%. Sentimiento positivo = 0. Sentimiento negativo = -6.

“Estos dias, no solo la comida es excesiva, la sobremesa es peor”

Importancia = 0,00378%. Sentimiento positivo = 0. Sentimiento negativo = -7.

Tuit realizado el día 3 de enero de 2014.

7.3.1.3. Análisis gramatical de los contenidos

A continuación vamos a realizar el análisis gramatical de todo el contenido generado por el usuario con código número 12:

- Total oraciones (tuits) = 104
- Longitud promedio de las oraciones (palabras) = 22,7
- Total palabras = 2.359
- Adjetivos más utilizados =

- buen (raíz), claro, mala, mejor, primero
- Adverbios más utilizados =
 - no, hoy, muy, ya, ahora, bien, más, tanto, mucho
- Conjunciones más utilizadas =
 - y, si, pero, como, o, que,
- Conjunciones subordinadas =
 - si, como, que, cuando, pues, aunque
- Verbos más utilizados =
 - es, puede, hace, hay, parece, ver, tiene,
- Palabras más frecuentes =
 - hoy, puede, jeje, año, hace, parece, resultados, ver
- Parejas de palabras más frecuentes:
 - Nueva, regulación
 - Pensaba, vender
 - Meses, tenemos
 - Sale, bien
 - Entre, todos
 - Me, gusta
 - Esta, mañana
 - Mucho, tiempo

Hemos incorporado el análisis gramatical al análisis de los sentimientos gracias al procedimiento de análisis que nos permite SETAL. Todo este proceso nos permite extraer las siguientes informaciones relevantes para su análisis:

- Palabras clave que el usuario utiliza con más frecuencia
- Combinaciones de palabras más frecuentes en su forma de expresarse
- Dentro de los *topics* observamos qué grupos de palabras son los más centrales en sus conversaciones.
- A través de la información reflejada en la columna: “Tot.Noms”, se pueden identificar aquellos amigos con los que se relaciona más a menudo.
- Qué tipo de verbos utiliza con mayor frecuencia.

- Uso de las palabras de contenido y palabras funcionales.
- Podemos determinar si es una persona en cuyas conversaciones predomina la polaridad positiva o negativa.
- Hallamos las oraciones más relevantes y centrales en sus conversaciones.

En definitiva y como habíamos explicado en capítulos anteriores, el análisis de un usuario o de un grupo de usuarios nos permite disponer de una información muy amplia y precisa que nos resulta básica para el conocimiento de lo que piensa y siente y nos muestra aquellos rasgos identificativos de su personalidad (preferencias, deseos y necesidades). Toda esta información es justamente la que necesita previamente un equipo de profesionales conocedores de la marca para proceder a elaborar los *insights* publicitarios.

7.3.1.4. Ejemplo de aplicación de la metodología “SETAL” para la elaboración de *insights*

A continuación presentamos un ejemplo de aplicación de la metodología “SETAL” para la elaboración de *insights*, a partir de una campaña publicitaria de una marca de moda (de ahora en adelante: Marca) que por razones de confidencialidad mantenemos su nombre en el anonimato.

Dicha Marca, es una reconocida empresa de moda cuya origen y sede está en Barcelona, con presencia internacional. Sus campañas publicitarias están orientadas a ser especialmente notorias, buscando un cierto punto de provocación, la cual no pasa desapercibida en las redes sociales. Éstas representan una importante fuente de información sobre las reacciones de sus consumidores, quienes las comparten tanto desde su polaridad positiva como negativa y generan un importante debate rico en contenidos de gran valor informativo para la Marca.

La campaña de publicidad del período comercial del “Día de la Madre” de 2014, generó un especial debate sobre aspectos éticos, con una gran repercusión tanto en medios de comunicación como en las redes sociales.

Así pues, la Marca consideró oportuno solicitar un análisis en profundidad para aprovechar todas estas manifestaciones de sus consumidores para obtener información sobre la percepción y posibles consecuencias de una campaña de un estilo tan “*agresivo*”.

Los objetivos fijados para este estudio eran, por un lado, la medición de los sentimientos asociados a los comentarios que sus consumidores habían expresado sobre la Marca, a raíz de su reacción ante la campaña de publicidad. Se pretendía medir el grado de aceptación o rechazo a través del análisis de la variable *Sentiment* y cuantificarla y por otro lado, realizar una prueba de la metodología SETAL para la elaboración de *insights* y aprovechar un motivo de reacción espontánea e intensa para identificar posibles *insights* para la Marca.

Aplicación de la metodología SETAL para la elaboración de *insights*:

Ámbito de estudio:

- Selección de palabras clave vinculadas con la marca y la campaña de publicidad desarrollada durante el período comercial del “Día de la Madre”.

Obtención de la información:

- Fuente: conversaciones de los usuarios en Twitter
- Período de análisis del 1 al 15 de mayo de 2014
- Tuits obtenidos: 66.392
- Usuarios seleccionados = 39.234

Filtros establecidos:

- Detección de idioma. Exclusión de todos exceptuando el castellano.
- Reiteración de mensajes (mensaje ya recogido por otros usuarios)
- Tuits inferiores a 6 palabras
- Que la fracción de “#”, “urls”,... respecto a la cantidad de palabras del texto sea inferior al 40%
- Tuits con predominio de palabras irrelevantes para el estudio.

Como resultado de los filtros, se obtienen las bases siguientes de análisis:

- Usuarios = 13.197 (33,6%)
- Tuits válidos = 18.196 (27,4%)

Análisis gramatical y de los sentimientos:

- Palabras analizadas = 360.325
- Sentimientos positivos (valores absolutos) = 19.185
- Sentimientos negativos (valores absolutos) = 36.606
- Ratio polaridad negativa / positiva = 1,91

Resumen:

La campaña de publicidad de la “Marca” con motivo del “Día de la Madre” no fue recibida de forma positiva por parte de sus usuarios ni del público en general. El ratio de tuits con polaridad negativa superó a los de polaridad positiva en una proporción de casi uno a dos.

Elaboración de *insights*:

Junto con toda la información (en parte eliminada por razones de confidencialidad) y con los textos de algunos tuits a título de ejemplo, un equipo de profesionales puede adecuar esta información al posicionamiento de la marca para poder obtener información relevante durante el proceso.

Comentario 1:

“Los anuncios de “Marca” anuncian, en ningún momento ropa. Pues yo sigo enamorada de los bolsos y abrigos de “Marca”. Aunque ahora da un poco de repelús entrar”.

Un comentario que denota una cierta lejanía del vínculo producto-marca.

Comentario 2:

“No es la apariencia, es la esencia. No es el dinero, es la educación. No es la ropa, es la clase”.

Es una frase de Coco Chanel, pero que debidamente procesada puede ser un eje de comunicación muy potente.

Comentario 3:

“La mejor marca de ropa que conozco se llama tu piel”

Es la letra de una canción bastante provocativa que aparece en el programa de radio “Ponte a prueba”. Observamos seguidamente dos oportunidades para posibles desarrollos en marketing y comunicación:

1. Producto para una línea de cosmética, colonia y similares
2. Producto de ropa de la “Marca” que protege la piel.

Como hemos ido explicando y mostrando a lo largo de esta tesis, la obtención de *insights* a través del estudio de los usuarios es una labor compleja y costosa que requiere tiempo y dinero. La herramienta SETAL de análisis de textos permite, de forma totalmente objetiva y no intrusiva, una vía alternativa para acceder a la información necesaria para que los profesionales puedan desarrollar su trabajo publicitario.

Para trabajar en la consecución de la información relevante, han sido especialmente útiles la utilización de los diferentes filtros para eliminar el ruido junto a la posibilidad de la jerarquización de los textos en función de los diversos parámetros y variables que hemos considerado prioritarios.

SETAL nos ha presentado la información más relevante de forma resumida y clara y con un nivel de profundidad que facilita los desarrollos comerciales a nivel de marca, marketing y comunicación.

En efecto, dicha información procesada y resumida nos permite identificar desde las tendencias, productos, servicios hasta establecer mejores líneas de comunicación y atender

con mayor eficacia a las necesidades de los clientes. Asimismo, podemos estudiar los valores de la competencia y muchos otros aspectos que hacen referencia a la comunicación de una empresa y de la marca que representa sus productos y sus valores.

En resumen, en este capítulo hemos presentado como bases metodológicas la técnica de la netnografía y la folcsonomía y hemos procedido propiamente a la investigación cuantitativa del análisis de contenido.

Gracias al análisis netnográfico hemos argumentado los beneficios que nos presenta la no intervención del investigador y que nos permite presentar un estudio sin sesgo.

Por otro lado, hemos utilizado las folcsonomías para identificar e interpretar las etiquetas que los usuarios han realizado de forma colaborativa en internet y que tienen su funcionamiento particular en Twitter.

Con la base metodológica de la investigación desplegada, hemos procedido a la aplicación del programa informático SETAL que nos ha permitido realizar el análisis gramatical y semántico y, en especial, el análisis de los sentimientos.

Avanzando en la aplicación de las opciones de SETAL hemos podido identificar qué contenido es más importante que otro y hemos podido acceder a él ordenándolo según diversas clasificaciones. Como hemos encontrado, el uso de según qué palabras funcionales por parte de los usuarios puede llegar a ser determinante para poder identificar la personalidad de un usuario. Si sumamos las palabras funcionales y el detallado análisis gramatical podemos identificar la “huella digital” que lo hace diferente a todos los demás.

En la subsección dedicada a la determinación de la muestra hemos presentado todo el proceso seguido para la selección del sector de actividad sobre el que desarrollar la investigación; hemos presentado la lista de palabras clave necesaria para identificar las conversaciones con contenido relacionado con nuestro ámbito de interés y argumentado los motivos por los cuales hemos considerado Twitter como nuestra fuente principal de información. Llegados a esta parte del capítulo, hemos descrito el proceso seguido para la obtención de la información con todas sus variables cuantitativas necesarias. Asimismo, hemos procedido a la aplicación de los filtros establecidos para separar la información relevante de la considerada “ruido” y

hemos elaborado un diccionario adhoc con más de 800 palabras clave relacionadas con el sector de actividad de interés presentando la validación estadística de la muestra final.

aEl objeto de estudio queda reflejado en el proceso de análisis de los contenidos según la identificación y la cuantificación de las horas del día, los días de la semana y los meses del año en que los usuarios se expresan con mayor o menor carga emotiva. Acto seguido, hemos identificado las oraciones más importantes y sus palabras clave, así como los usuarios más relevantes y los criterios para elegirlos.

Para presentar los resultados se ha optado por la segmentación de los usuarios según *topics* (en función del uso similar del uso de determinados grupos de palabras homogéneas), creyendo que de este modo aislamos satisfactoriamente los atributos y características fundamentales tanto de los contenidos como de los usuarios. Posteriormente, basándonos en esta segmentación, hemos podido establecer relaciones entre *topics*, grupos de usuarios y tipología de contenidos. Ello nos ha reportado información de alto valor para profundizar en los aspectos que un profesional de la comunicación publicitaria puede necesitar como material para la elaboración de *insights*. En la fase final del proceso, hemos podido vincular la información alcanzada a través de los *topics* con la presencia de sentimientos positivos y negativos y hemos finalizado con un estudio detallado y profundo del análisis gramatical de cada usuario.

Como cierre del capítulo, hemos presentado un ejemplo de aplicación de la metodología SETAL de elaboración de *insights*, realizado por una marca de moda con sede en Barcelona y presencia internacional que por razones de confidencialidad no podemos revelar ni su nombre ni parte del contenido del trabajo. Sin embargo, sí hemos podido valorar, con independencia del resultado, que los datos y la información obtenida con el método sí es válida para desarrollar el trabajo publicitario mediante el hallazgo de *insights* publicitarios.

8. INVESTIGACIÓN CUALITATIVA

8. INVESTIGACIÓN CUALITATIVA

La aplicación de SETAL para realizar el análisis cuantitativo expuesto en el capítulo 4 supone el núcleo principal de esta tesis y el trabajo con los resultados obtenidos nos ha permitido desarrollar una metodología para la elaboración de *insights*. La decisión de añadir una investigación cualitativa se tomó con el objetivo de completar al máximo la información total a tener en cuenta para la toma de decisiones, pero no desde el punto de vista de los usuarios sino desde el punto de vista de las personas responsables de las empresas que pueden aplicar nuestro método de investigación en sus respectivas áreas de comunicación y publicidad.

Con este propósito preparamos una serie de entrevistas en profundidad a diversos directores/as de los departamentos de comunicación digital de las empresas más representativas del sector bancario, el sector de actividad sobre el que habíamos realizado previamente la investigación empírica.

Nuestro interés al abordar las entrevistas radicaba en la voluntad de ampliar nuestro conocimiento acerca de la presencia del sector bancario en las redes sociales, la importancia que sus directores le atribuían a su uso, los métodos de monitorización de las comunicaciones en las redes y también si las estaban utilizando de algún modo para obtener información relevante para la elaboración de *insights* con la inclusión de la variable *Sentiment*.

Consideramos que la mejor manera de obtener toda esta información era mediante la realización de entrevistas en profundidad individuales y estructuradas en base a una serie de preguntas.

Las entrevistas fueron grabadas y se solicitó la autorización para difundirlas exclusivamente en el ámbito académico. Fueron realizadas durante el período comprendido entre julio y noviembre de 2014.

Siempre con la voluntad de centrar nuestro foco de atención en el campo publicitario y en los procesos de elaboración de campañas publicitarias, consideramos conveniente ampliar las entrevistas más allá del sector bancario, con lo que realizamos diversas entrevistas a directores/as de los departamentos mencionados de empresas de gran consumo, concretamente pertenecientes al sector de alimentación. Esta decisión fue tomada por dos motivos: por un lado, debido a su fuerte actividad tanto en el campo de la comunicación como en las redes sociales y, por el otro, porque hemos trabajado a lo largo de esta tesis con diversos casos reales de empresas de este sector.

De hecho, como pudimos ver en la sección 2.1 de los métodos de investigación aplicadas a la comunicación, Nestlé es un anunciante que ha desarrollado su propio método para la elaboración de *insights*, razón que nos llevó a entrevistar a la persona responsable del área de la comunicación digital de la empresa. Además de Nestlé, entrevistamos a la persona con el mismo cargo de la empresa Danone.

Con la información recibida, nos pareció interesante analizar y comparar los puntos que mantenían en común y las divergencias respecto a las preguntas planteadas, aunque como empresas puedan tener filosofías y prioridades diferentes a todos los niveles. Los puntos de coincidencia se deben principalmente a que ambas empresas se dirigen a un mismo público objetivo.

8.1. Metodología de la investigación cualitativa

Nuestro modo para seleccionar a las personas que necesitábamos entrevistar fue a partir de la selección de las principales empresas del sector bancario de España y que representamos en este gráfico con datos de El País:



Figura 52: Ranking de Bancos Españoles por activos. (Catalán, 2015)^[445]

A partir de la clasificación de los bancos españoles, nos pusimos en contacto con ellos para solicitar una entrevista personal con los responsables de las áreas de gestión de las redes sociales y/o del área de comunicación digital.

Pudimos obtener entrevistas con dichos responsables de los bancos siguientes: CaixaBank, Banco Sabadell-Atlántico, BBVA y Catalunya Caixa (entrevista realizada previamente a su absorción por el BBVA). A esta serie de entrevistas añadimos las de Nestlé y Danone y alcanzamos un total de seis entrevistas con catorce preguntas en cada una de ellas.

Con la finalidad que hemos descrito, tuvimos presente estos tres puntos en la elaboración de las preguntas:

1. Conocer la presencia de las marcas en las redes sociales y conocer cómo organizan su estrategia de comunicación.
2. Obtener información sobre las métricas que consideran más relevantes y las herramientas que utilizan para su medición.
3. Determinar el grado de interés que tiene la variable *Sentiment* en sus análisis .

Ampliamos a continuación cada uno de los puntos considerados:

1.- Conocer la presencia de las marcas en las redes sociales y su organización

En primer lugar, quisimos conocer cuáles son las principales redes sociales de trabajo tanto para los bancos como para las empresas de gran consumo. En segundo lugar, nos fijamos en las razones por las que consideran la necesidad de estar presentes en dichas redes sociales y qué tipología de organización u organizaciones han seleccionado teniendo en cuenta su gran tamaño como empresa y la importante variedad de productos y servicios que ofrecen. En este sentido, nos resultaba especialmente relevante el saber si estas empresas subcontratan esta gestión o si la gestionan internamente y, en caso de realizarlo dentro de su propia estructura, nos interesaba saber si la organizaban por departamentos u optaban por una separación por productos.

2.- Información sobre las métricas más relevantes y las herramientas que utilizan para obtenerlas

Conscientes de la gran cantidad de datos que se generan en las redes sociales por parte de los usuarios, ya sea como respuesta a las campañas de los anunciantes como mostrando y compartiendo sus opiniones y otros contenidos, consideramos de especial interés conocer qué información es la que el sector bancario y el de gran consumo monitorizan y analizan y por qué motivos. Del mismo modo, nos interesaba saber qué herramientas de resumen y análisis

utilizan, la frecuencia de realización de informes y, en especial, el uso posterior que realizan de esta información.

3.- Grado de interés de la variable Sentimient en sus análisis

Como hemos dejado patente, consideramos fundamental la variable *Sentimient* para la elaboración de *insights*. En concreto, queríamos saber si estos sectores monitorizan esta variable, cómo la monitorizan (en el sentido del grado de granularidad de la información) y con qué problemas se han encontrado. Asimismo, nos interesaba conocer el grado de fiabilidad en sus trabajos y, finalmente, si utilizaban esta información para la toma de decisiones y, en concreto, si la tenían en cuenta para la elaboración de *insights*.

8.1.1. La entrevista en profundidad

8.1.1.1. Modelo de la entrevista en profundidad

A continuación presentamos el modelo en entrevista individual utilizado para la investigación cualitativa.

Investigación cualitativa de la tesis doctoral sobre del doctorando Xavier Duran.

Fecha: Barcelona a, xx de xxxx de 2014

Entrevista al profesional

- Nombre :
- Apellidos :
- Empresa :
- Cargo :
-

Introducción

Los *insights* de los consumidores han sido y son una fuente necesaria de información de alto valor para las empresas y para la elaboración de campañas de comunicación publicitaria para las agencias de publicidad.

Los métodos más comunes para su obtención han sido las entrevistas, reuniones y encuestas, entre otros. Debido a la dinámica del propio grupo entrevistado y a la presencia del entrevistador, entre otros factores, se produce un sesgo en las manifestaciones de los participantes.

Por otro lado, los usuarios y consumidores (todos nosotros) utilizamos y manifestamos libremente nuestras ideas, sensaciones, emociones, sentimientos y opiniones en las redes

sociales para compartirlas con familiares, amigos, otros profesionales del sector o personas afines a nuestros intereses. Es así que las redes sociales reflejan millones de comunicaciones, desde tuits breves hasta artículos extensos o todo tipo de opiniones acerca de productos, marcas o servicios. Toda la información compartida y a la que podamos acceder de forma legítima puede resultar de utilidad para el trabajo de comunicación de las empresas. Sin embargo, debido a su enorme abundancia (Big Data), necesitamos herramientas que elijan, resuman y analicen el contenido que nos interesa y que nos ofrezcan un resultado comprensible y manejable para que podamos utilizarlo.

Esta entrevista pretende profundizar en las herramientas y el uso que las empresas están haciendo actualmente de esta posible fuente de información.

Autorización

Quisiera pedirle su autorización para utilizar la información que usted me facilite durante esta entrevista para utilizarla en el estudio cualitativo relativo a la tesis doctoral antes mencionada y en las publicaciones que de este estudio se deriven.

Inicio de la entrevista. Preguntas:

1. ¿Cuál es el papel de las redes sociales en su empresa?
2. ¿Por qué su empresa está presente en las redes sociales?
3. ¿Cuáles son los principales beneficios que espera conseguir con su presencia en las redes sociales?
4. ¿En qué redes sociales su empresa está presente (tiene una o varias cuentas creadas)?
 Facebook Twitter Youtube LinkedIn Google Instagram
 Pinterest Tuenti _____
5. ¿Cuál de las redes sociales considera más importante para su empresa? (mencionar solo una)

Facebook Twitter Youtube LinkedIn Google Instagram
 Pinterest Tuenti _____

6. ¿Cómo se realiza la gestión de las redes sociales? (internamente, subcontratada, parcialmente subcontratada)
7. Dentro de la empresa, ¿qué personas son las que gestionan las redes sociales?
8. ¿De qué departamento depende este equipo de personas?
9. ¿Monitorizan la presencia en las redes sociales?
10. ¿Qué herramientas de monitorización utilizan para el seguimiento de su presencia en las redes sociales?
11. ¿Con qué frecuencia emiten informes de seguimiento de la actividad en las redes sociales?
12. ¿Qué métricas consideran como las más importantes a monitorizar?
13. ¿Han utilizado alguna vez las herramientas de monitorización para identificar posibles *insights*?
14. ¿Monitorizan la métrica *sentiment* (valora los comentarios positivos, negativos o neutros) que realizan sus consumidores?

8.1.1.2. Resultado de las entrevistas en profundidad

En esta sección resumiremos los resultados obtenidos fruto de las entrevistas realizadas. Con ello mostraremos el cumplimiento de los objetivos principales de esta fase de la investigación cualitativa.

Pregunta número 1: *¿Cuál es el papel de las redes sociales en su empresa?*

El sector bancario está utilizando las redes sociales especialmente como apoyo al servicio de atención al cliente, el marketing, la comunicación de su actividad y también para comunicarse con los medios de comunicación (prensa, periodistas, etc.).

Las redes sociales se han convertido para este sector en un importante canal de comunicación que les permite transmitir el posicionamiento deseado, mantener el diálogo y la relación directa con los consumidores y obtener con ello información que les permita anticiparse a sus necesidades y poder ofrecerles soluciones a medida de sus necesidades. En este sentido, las redes sociales son un canal que las empresas tienen cuenta y suman a los otros canales para fortalecer su estrategia multicanal. Las entrevistas muestran el valor que otorgan las empresas a la posibilidad de proximidad que ofrecen las redes sociales en cuanto a la proximidad con los clientes.

Pregunta número 2: *¿Por qué su empresa está presente en las redes sociales?*

Dentro de un proceso de transformación hacia un entorno digital, los bancos han desarrollado una gran cantidad de servicios vinculados a la banca online para facilitar los servicios a sus clientes. Como ejemplo paradigmático, el servicio “eLaCaixa” ha desarrollado redes sociales propias y exclusivas para segmentos específicos de clientes.

Las redes sociales se consolidan como los medios de comunicación ideales para el desarrollo de relaciones directas con los clientes a largo plazo y destacan por la seguridad y confianza que las empresas transmiten al cliente con el hecho de poder atenderles de forma rápida y dándoles la confianza de saber que hay un equipo de personas que permanecen su servicio.

A nivel publicitario, las redes sociales alcanzan a millones de personas con lo que, para realizar campañas pueden alcanzar un volumen antes solo atribuible a los medios de comunicación de masas. Ya sea por la vía de la atención al cliente o por por la vía publicitaria, las redes sociales se consideran óptimas para fidelizar a clientes existentes y para captar a nuevos clientes potenciales.

Pregunta número 3: *¿Cuáles son los principales beneficios que espera conseguir con su presencia en las redes sociales?*

Como indicaban las preguntas anteriores, los puntos más comunes en las respuestas a esta pregunta (tanto en las empresas del sector bancario como en las de gran consumo) se refieren a los beneficios que ofrecen las redes sociales como medios interactivos de diálogo con los clientes y en la reconfortante sensación de proximidad que sienten los usuarios si se utilizan con esta finalidad. Además, esta cercanía permite tanto la fidelización de clientes como la atracción de nuevos públicos y es una excelente vía para conocer sus necesidades e inquietudes.

Pregunta número 4: *¿En qué redes sociales su empresa está presente (tiene una o varias cuentas creadas)?*

Dado que las empresas seleccionadas son grandes anunciantes manifestaron estar presentes en casi todas las redes sociales, bien de forma intensiva como en Facebook y Twitter como a nivel de prueba para descubrir su evolución, como en el caso de Instagram. Solo una de las empresas entrevistadas está presente en Tuenti.

Pregunta número 5: *¿Cuál de las redes sociales considera más importante para su empresa? (mencionar sólo una)*

Facebook es la red social más utilizada por todas las empresas entrevistadas a excepción de una de ellas que manifestó utilizar más Twitter que Facebook.

La razón principal que determina la importancia es el alcance del público. En el caso de Nestlé, destaca su liderazgo por la publicación de vídeos sobre recetas de cocina.

Pregunta número 6: *¿Cómo se realiza la gestión de las redes sociales? (internamente, subcontratada, parcialmente subcontratada)*

Se observa una evolución hacia la gestión interna de un sector tan atento a la relación directa con los clientes. Sin embargo, los modelos de organización son sustancialmente diferentes.

Hay empresas que siguen manteniendo una organización mixta, de tal modo que el grueso de la relación directa con los consumidores se realiza internamente y la creación de contenidos, como las promociones, se subcontrata a agencias externas.

El modelo de organización interna difiere según la empresa. Podemos encontrar desde un departamento de comunicación corporativa que coordina toda la comunicación (incluso a nivel internacional) hasta un modelo de gestión totalmente transversal entre departamentos o, en otro caso, una gestión global a nivel de marca.

Todas las empresas consideran que estos departamentos tienen un rol muy estratégico y los *community managers* son las figuras responsables de mantener la relación entre el contenido, la línea editorial y la relación con el consumidor.

Pregunta número 7: *Dentro de la empresa, ¿qué personas son las que gestionan las redes sociales?*

Los responsables de la gestión de las redes sociales se determinan en función de los diferentes sistemas de organización. Según manifiestan los entrevistados, los responsables de las redes sociales pueden formar parte de una unidad centralizada en el departamento de comunicación corporativa pero posteriormente es posible que la gestión diaria recaiga en diferentes áreas de negocio. Por otro lado, nos hemos encontrado con una gestión transversal entre los diferentes departamentos en función del contenido de los productos sin que sean los mismos responsables de los productos los que desarrollen la parte de la comunicación. En

conclusión, hay empresas que no tienen un responsable concreto de redes sociales mientras que otras reconocen la importancia de la figura de los *community managers* dentro de su equipo de comunicación digital.

Pregunta número 8: *¿De qué departamento depende este equipo de personas?* Dado que las respuestas son cortas, realizaremos una transcripción más literal de las respuestas obtenidas:

- Dos empresas están organizadas a través de su departamento de marketing.
- Dos empresas están organizadas por departamentos (organizados por líneas de productos).
- Encontramos otra empresa que tiene el equipo dividido en cuatro departamentos, siendo el de atención al cliente el que tiene más personal asignado.
- Otra empresa lo separa entre el Departamento de comunicación y el de marca.

Normalmente, estos departamentos dependen a su vez del departamento de marketing y suelen gestionar las redes por productos. Hay empresas que han declarado tener un responsable diferente de cada red social.

Pregunta número 9: *¿Monitorizan la presencia en las redes sociales?*

Todas las empresas monitorizan su presencia en las redes sociales, tanto aspectos cuantitativos (alcance, menciones, entre otros) como aspectos cualitativos.

Cabe destacar el espacio físico que posee eLaCaixa especialmente adaptado para la monitorización de la presencia de la empresa en las redes sociales. eLaCaixa cuenta con una sala con multitud de pantallas que muestran el seguimiento permanente de diferentes métricas o acciones en diferentes redes sociales, tanto las redes como Facebook o Twitter como las que han desarrollado de manera interna.

Por otro lado, resulta interesante el proyecto desarrollado por Nestlé a nivel internacional llamado *Digital Acceleration Team*. El proyecto está centralizado en Vevey (Suiza) y se dedica a la monitorización de la presencia de la marca y de los comentarios de los clientes en todas las redes sociales. Según manifiesta Nestlé, este proyecto es clave para entender bien al consumidor y estar más cerca de él. La tecnología es en este caso una herramienta que permite conectar al consumidor con las marcas.

Ambos proyectos demuestran la función estratégica que tienen las marcas para identificar la información relevante en un entorno *Big Data*.

En este sentido, Nestlé añadió en respuesta a esta pregunta el uso de la monitorización para la identificación de *insights* con la finalidad de innovar y poder adecuar así sus campañas de publicidad y renovar la generación de contenidos.

Pregunta número 10: *¿Qué herramientas de monitorización utilizan para el seguimiento de su presencia en las redes sociales?*

De las seis empresas entrevistadas, tres mencionan la herramienta de pago Radian6. También manifiestan que, al ser un sector relativamente nuevo, las empresas han ido probando diferentes herramientas a medida que han ido apareciendo y que, por lo tanto, tienen una variación importante de soluciones y herramientas.

La mayoría combina diferentes herramientas, tanto gratuitas (principalmente de Google, como Google Alerts) como de pago.

Es importante destacar que las empresas remarcan el hecho de que, a pesar de que estas herramientas resultan imprescindibles para resumir y analizar la información facilitando así la labor de los analistas, el papel fundamental recae en las personas que interpretarán los resultados obtenidos y tomarán con ellos las decisiones pertinentes.

Pregunta número 11: *¿Con qué frecuencia emiten informes de seguimiento de la actividad en las redes sociales?*

Esta pregunta genera una respuesta coincidente y unánime. Todas las empresas realizan un seguimiento diario de su presencia en las redes sociales.

Las empresas manifiestan la existencia de diferentes categorías de informes determinados por su frecuencia (alertas diarias, informes semanales, mensuales) o en función de la acción o de un tema de interés.

Pregunta número 12: *¿Qué métricas consideran como las más importantes a monitorizar?*

Principalmente, valoran las métricas cuantitativas: desde las que permiten eliminar el gran volumen de “ruido” a las que permiten determinar las audiencias, coberturas y la actividad en general en las diversas redes sociales. También valoran en especial la medición del tráfico (visitas) que generan las redes sociales hacia el sitio web.

En cuanto a las métricas cuantitativas, se interesan por la identificación de las polaridades (negativa, positiva o neutra) de los comentarios de los usuarios, el volumen de interactividad entre las marcas y los usuarios que generan los contenidos, el análisis de tendencias y la identificación de *clusters*.

Pregunta número 13: *¿Han utilizado alguna vez las herramientas de monitorización para identificar posibles insights?*

De una manera u otra, todas las empresas entrevistadas manifiestan que utilizan las redes sociales para obtener información relevante que sirve para la elaboración de *insights*. Desde una respuesta clara como en el caso de Nestlé hasta respuestas que realizan un recorrido descriptivo de cómo lo llevan a cabo.

Por ejemplo, eLaCaixa informa a su red social privada con “Inspiran’s”. Esta vía representa una fuente de obtención sistematizada de información de los clientes pidiéndoles a los propios clientes información acerca de aquello que necesitan o que les gustaría que su banco hiciese por ellos. Gracias a este mecanismo de comunicación han surgido productos como la Cyber Tarjeta o el pack básico de alta de cliente online, entre otros.

El BBVA realiza el análisis de los comentarios positivos y/o neutros para la elaboración posterior de mensajes, ideas e incluso productos. Como comentábamos, en el caso de Nestlé, la empresa considera importante saber si se está conectando con el consumidor a partir de los *insights* que se han identificado previamente y si se han adecuado correctamente los mensajes en función de los *insights* obtenidos y las tipologías de los consumidores detectadas.

Pregunta número 14: *¿Monitorizan la métrica sentiment (valora los comentarios positivos, negativos o neutros) que realizan sus consumidores?*

La pregunta final de la entrevista ofrece cierta disparidad en las respuestas. Desde empresas que manifiestan darle mucha importancia hasta empresas que manifiestan todo lo contrario.

Como opinión general, las empresas consideran que las herramientas actuales tienen que evolucionar para llegar a ser capaces de captar todos los matices del lenguaje escrito. Como apuntamos en esta tesis, ellas también constatan que por ahora no son capaces de medir correctamente ni el uso de la ironía ni de los dobles sentidos; por ahora, su principal utilidad reside en su capacidad para resumir y destacar lo más importante dentro del “ruido” existente y en permitirles centrar la atención en los temas que más interés despierta en los consumidores.

Las respuestas también destacan el papel clave del analista que interpreta las informaciones obtenidas y las contextualiza antes de la toma de decisiones.

A modo de resumen, la realización de las entrevistas para esta parte de la investigación cualitativa ha complementado con información relevante los contenidos obtenidos previamente mediante la información académica y la investigación. Al plantear la posibilidad de aplicación en el ámbito profesional de nuestra investigación era preponderante realizar un acercamiento profundo a algunos de los principales actores de nuestro campo de investigación.

Empresas líderes como La Caixa nos hablaron de proyectos concretos (eLaCaixa) especialmente diseñados para funcionar en un entorno *online* y conectado a través de las redes sociales. Nestlé nos presentó su centro internacional de monitorización de las redes sociales con su llamado *Digital Acceleration Team*.

Mediante estas entrevistas hemos conseguido detectar los principales beneficios por los cuales estas empresas están presentes en las redes sociales y cómo se organizan para funcionar dentro de ellas. Afirmamos que las redes sociales ya se consideran medios de

comunicación masiva y que se utilizan tanto desde el punto de vista publicitario como el de la relación con los clientes, medios y otras entidades. De este modo, constatamos el cambio de paradigma en el ámbito de la comunicación de masas, que va virando hacia un modelo basado en la comunicación directa, interactiva y personal entre las marcas y sus consumidores. Las mejoras para la comunicación de las empresas son notables y se han mencionado en repetidas ocasiones: mejora el conocimiento de sus clientes, pueden adecuar mejor sus campañas de comunicación y también sus servicios de atención al cliente, así como valorar con mayor efectividad la receptividad de los productos y servicios que lanzan al mercado.

Casi sin excepción, la red social más utilizada es Facebook, seguida de Twitter. Como grandes anunciantes, están presentes en prácticamente todas las redes sociales aunque en alguna de ellas su presencia es básicamente para valorar su evolución y si merece la pena mantener una continuidad.

Respecto a la gestión de las redes sociales, la tendencia es a realizarla de manera interna, aunque con notables diferencias en cuanto a su gestión.

Las métricas más monitorizadas son las cuantitativas aunque con un interés creciente por las variables cualitativas. En cuanto a la variable *Sentiment*, los bancos le dan menos relevancia debido a que prefieren mantener una relación más cauta con sus clientes. Las herramientas les permiten detectar con suma facilidad los temas de interés y de actualidad de sus clientes, así como de la evolución de los mismos.

Respecto a la elaboración de *insights* haciendo uso de la información obtenida a través de las redes sociales, hemos podido constatar que todas las empresas, de una manera u otra, han aprovechado los datos para esta finalidad.

En cuanto al uso y la atención prestada a la variable *Sentiment*, la respuesta ha sido dispar. Mientras que hay empresas que consideran que las herramientas actuales no son capaces de medirla de forma adecuada, otras empresas la analizan y consideran que la información obtenida sí es relevante para sus comunicaciones. Sin duda, las empresas reivindican la figura

del analista como persona clave para convertir la información en comunicaciones e ideas de éxito.

La variable *Sentimient* se considera importante para la identificación de posibles situaciones de crisis, puesto que las polaridades se acentúan y pueden ser identificadas con rapidez.

En todas las entrevistas realizadas existe un consenso acerca de que el análisis continuado de la esta variable marca una tendencia interesante y aporta un valioso ángulo de visión global sobre la percepción de la imagen de marca.

Respecto a las perspectivas de futuro del sector de la comunicación en redes sociales, las empresas entrevistadas han considerado que, de cara a seguir mejorando sus comunicaciones, las herramientas de resumen y análisis evolucionarán y mejorarán y que, en este sentido, la variable *Sentimient* se medirá con mayor precisión, siendo su papel cada vez más central en sus informes y en su toma de decisiones.

9. CONCLUSIONES

9. CONCLUSIONES

En este capítulo presentaremos las conclusiones obtenidas mediante el trabajo de investigación desarrollado. Revisaremos los objetivos y las hipótesis planteadas al inicio de este trabajo de investigación y las contrastaremos con los resultados obtenidos con el fin de validarlos o refutarlos.

Recordaremos, uno por uno, los tres objetivos de la tesis y las hipótesis de las que partimos y realizaremos los comentarios pertinentes valorando el resultado de nuestra investigación:

Objetivo 1. Demostrar que la selección y análisis de la información que se puede obtener en las redes sociales, en concreto a través de la plataforma Twitter, es una fuente de información que puede ser utilizada para la elaboración de *insights* publicitarios.

Hipótesis 1. El algoritmo desarrollado específicamente para esta tesis doctoral, denominado SETAL acrónimo de *Sentiment Text Analysis*, que permite la obtención de información relevante del usuario y de la nueva metodología creada con el mismo nombre, se pueden elaborar *insights* publicitarios.

Con la aplicación del algoritmo desarrollado a través de la herramienta SETAL y de la metodología del mismo nombre, logramos encontrar información relevante que expusimos en la segunda parte de la tesis, en el capítulo sexto. En efecto, por una lado la herramienta SETAL nos permite obtener una gran cantidad de información procedente del resumen y análisis de textos, encontrar los temas de interés de los usuarios, la información que comparten, las razones que los llevan a hacerlo, las personas con las que la comparten y los sentimientos que expresan en sus conversaciones a través de la plataforma Twitter. SETAL realiza un análisis gramatical detallado de todos los textos, agrupando y jerarquizando las palabras, oraciones, textos, *topics*, usuarios y sentimientos más *importantes*, que como ya

comentamos no son los más repetidos ni más frecuentes, sino aquellos que ocupan una posición de centralidad en el grafo de textos y su distribución estadística.

Como hemos visto en el estudio de los métodos tanto tradicionales como desarrollados por algunas empresas, para la elaboración de *insights*, la primera e imprescindible fase, consiste en la obtención de información relevante sobre el consumidor y su relación con la marca y producto o servicio de interés, factor que ha quedado claramente demostrado.

Por otro lado, hemos desarrollado una nueva metodología para la elaboración de *insights* (que volvemos a detallar en estas conclusiones más adelante) de la cual consideramos oportuno destacar sus elementos más novedosos y diferenciales a los otros métodos analizados y que corresponden a su mayor nivel de representatividad de la muestra, dado que se analizan cientos o miles de usuarios en lugar de unos pocos. SETAL elimina el sesgo dado que el entrevistador o investigador no interviene en las conversaciones que los usuarios están realizando de forma totalmente libre y espontánea a través de Twitter y finalmente, toda la información es agrupada y jerarquizada por los criterios de *importancia*, antes mencionados, de forma que reduce, sino elimina el factor *¡Ajá!* o *descubrimiento*, fruto de una iluminación súbita del investigador para poner en valor una metodología basada más en la deducción que en la intuición. En definitiva, más científica.

Aplicamos la metodología SETAL para la elaboración de *insights* en un estudio práctico para una empresa de moda con sede en Barcelona y presencia internacional. Es un gran anunciante que realiza grandes campañas de comunicación a través de todos los medios y en especial, la televisión. El tono desenfadado y hasta cierto punto provocador del tono de su comunicación tiene inmediatos efectos en las redes sociales, donde los comentarios representan expresiones fieles de los sentimientos que dichas campañas generan en la audiencia. Aplicamos la nueva metodología SETAL para el análisis de todos aquellos usuarios que, durante un período determinado de tiempo, realizaron comentarios en Twitter sobre la campaña publicitaria que desarrolló la MARCA durante el período promocional del “Día de la Madre”. SETAL pudo medir y cuantificar los sentimientos de rechazo a la campaña, identificar las palabras más *importantes* del discurso, en definitiva, se obtuvo un análisis gramatical, semántico y de los sentimientos expresados, con el máximo nivel de detalle y una vez resumidos y jerarquizados los puntos de interés, se pudo realizar un análisis experto que permitió identificar discursos que permiten la obtención de *insights* y que se sugirieron en el informe que se presentó a la

MARCA. Este ejercicio, permitió poner en práctica la nueva metodología y validar su eficacia.

En este sentido, validamos la hipótesis inicial del objetivo 1 que asumía que Twitter es un canal útil y pertinente y justificamos su valor para la obtención de información verdadera y fidedigna de los usuarios y que, con el análisis de dicha información y la aplicación de la nueva metodología SETAL, podemos obtener *insights* publicitarios.

Objetivo 2. Demostrar que, a través de Twitter y, por extensión, de las redes sociales, se puede obtener información que también sirva para los profesionales de la comunicación publicitaria y los departamentos de marketing de las empresas.

Hipótesis 2. Las herramientas de resumen automático que utilizaremos ofrecen unos resultados que son útiles, además de para los publicitarios, para otros departamentos de la empresa.

Como hemos visto, las opiniones de los consumidores y las conversaciones que realizan en las redes sociales se pueden resumir automáticamente mediante SETAL y pueden ser utilizadas para mejorar aspectos como la cercanía con los clientes o las comunicaciones concretas para cada producto lanzado al mercado.

Fruto de la labor de investigación a través de la herramienta SETAL y de la creación de la nueva metodología del mismo nombre para la elaboración de *insights*, hemos obtenido información como la medición o cuantificación de los sentimientos que los usuarios expresan a través de sus tuits, en función de las horas del día, los días de la semana o los meses del año. Asimismo, se obtiene información relevante sobre de qué hablan y cómo se expresan los usuarios analizados. Como hemos comentado, SETAL jerarquiza varias variables de análisis como las palabras, oraciones, sentimientos, por su *importancia*. Pero nuestro nivel de análisis puede ir más allá, dado que puede identificar los sentimientos que expresan los usuarios mediante la cuantificación numérica de la polaridad positiva, negativa o neutra de sus sentimientos. Es decir, no solamente identifica si son positivos o negativos sino que los cuantifica mediante valores numéricos que permite jerarquizarlos.

La cantidad y calidad de la información obtenida permite obtener un profundo conocimiento del consumidor que puede ser utilizado tanto por los departamentos de marketing como de comunicación en beneficio mutuo. Consideramos que, todo este conocimiento del usuario o consumidor, tiene grandes aplicaciones tanto para los departamentos de marketing como de comunicación de los anunciantes, organismos públicos, entre otros.

Estamos ante una nueva era de la publicidad científica, una creciente necesidad de poder gestionar el *Big Data*, en beneficio de una aproximación más directa entre las marcas y los consumidores y de una respuesta más adecuada de las empresas ante las necesidades de sus consumidores, lo que en definitiva, es la esencia del marketing y de una de sus herramientas más valiosas, la comunicación, que cada vez tiende a ser más personal y personalizada.

Para la confirmación de esta segunda hipótesis, la investigación cualitativa ha contribuido a reforzar su validación, dado que los departamentos de marketing, comunicación y gestión de las redes sociales de las empresas entrevistadas, nos han manifestado su creciente interés y desarrollo de estas aplicaciones. En este sentido, la empresa Nestlé ha desarrollado un programa específico, a nivel internacional con sede en Suiza, de monitorización y seguimiento de todas sus campañas, denominado: Digital Acceleration Team (DAT) consistente en realizar un seguimiento constante de toda la actividad y presencia de Nestlé en los medios de comunicación y en especial en los medios sociales. El DAT obtiene la información que distribuye a los departamentos de marketing por lo que se refiere a producto, a los departamentos de atención al cliente para mejorar su gestión y a los *community managers* para adecuar su discurso y saber sobre qué temas focalizar sus esfuerzos.

Si bien es cierto que este caso es un ejemplo perfecto de aplicación de la gestión del *Big Data*, es todavía una iniciativa poco común y estamos seguros que cada vez habrá más empresas que integren esta información y gestión dentro de su actividad cotidiana.

Así pues, validamos la hipótesis inicial del objetivo 2 sobre la importancia, utilidad y conveniencia de la utilización y gestión de estas informaciones por parte de los departamentos de marketing y comunicación de las empresas.

Objetivo 3. Analizar el estado del estudio, análisis y utilización de la variable *Sentiment* de los consumidores en las estrategias y acciones de comunicación de los anunciantes.

Hipótesis 3. La variable *Sentiment* se tiene cada vez más en cuenta para la elaboración de las estrategias de marketing y comunicación de las empresas.

Con la investigación cuantitativa demostramos la posibilidad de detectar y medir la variable *Sentiment* en las comunicaciones de los usuarios. Además, presentamos un caso real en el que se tuvo especialmente en cuenta esta variable. Sin embargo, los resultados de la investigación cualitativa mostraron una diversidad de opiniones desde el sector profesional respecto a la importancia de dicha variable en el campo publicitario. Si bien confirmaron de forma unánime tener interés hacia la información que proporcionan los sentimientos compartidos por los usuarios, algunos mostraron sus reservas a la hora de utilizarlos debido a que los sistemas de análisis todavía no han desarrollado un nivel de precisión suficiente como para detectar matices tales como la ironía o el doble sentido. De todos modos, otras empresas sí afirmaron tener dicha variable en consideración tanto para sus comunicaciones como para la realización de trabajos publicitarios. Al ser preguntados al respecto de cómo veían esta situación en el futuro, su visión era optimista y todos apostaban por una mejora constante de las herramientas de análisis y la incorporación de nuevos algoritmos capaces de medir cada vez información más compleja (metáforas, ironía, humor, etc.) respecto a la variable *Sentiment*. Sin embargo, todas las empresas mencionaban la importancia del conocimiento y las capacidades de los investigadores para interpretar los resultados obtenidos de manera satisfactoria. En consecuencia, no podemos validar esta hipótesis en su totalidad y debemos hacerlo con los matices que acabamos de señalar, con expectativas de notables mejoras en un futuro cercano.

La principal contribución de esta tesis ha sido, más allá de los objetivos iniciales planteados, el desarrollo de una metodología para la elaboración de *insights* a través del RAT y el SA realizados por el programa SETAL. Con el desarrollo del capítulo cuarto logramos estructurar una serie de fases cuya culminación es recoger una serie de datos con una calidad pareja a la que necesitan los profesionales de la comunicación publicitaria y que anteriormente solían obtener mediante el uso de los métodos llamados “tradicionales” (con mayor coste, limitaciones de la muestra y diversos sesgos que hemos definido con anterioridad). Consideramos que con una herramienta como SETAL y una metodología como la presentada es posible situarse en unas condiciones óptimas para elaborar *insights*. De hecho, con el método que hemos presentado consideramos que se logra reducir la dependencia de la intuición, mucho más volátil e intangible, y aportamos una serie de información que apuesta por una vía más científica y sólida.

La aplicación de nuestra metodología partió de la investigación que realizamos en profundidad en la literatura académica, que nos mostró los métodos habituales utilizados para la elaboración de *insights*. Dichos métodos tienen los siguientes puntos en común:

- Identificación del problema y definición del ámbito de interés.
- Obtención de la información: palabras clave, de uso más frecuente, según su relevancia, detección de pensamientos e ideas importantes, emociones. Junto con el estudio del consumidor: gustos, necesidades, deseos o inquietudes.
- Análisis de la información y agrupación de las observaciones en distintas clasificaciones.
- Extracción de los *insights* gracias al momento de descubrimiento (*jajá!*). El hallazgo de *insights*, la ausencia de ellos o la eficacia si finalmente se consigue extraerlos dependerá en gran medida de las capacidades interpretativas de los investigadores y publicistas que manejen los datos resultantes de la metodología presentada.

El método “SETAL” que proponemos desarrolla las fases siguientes, que coinciden con las técnicas mencionadas con anterioridad y en otro tipo de metodologías:

- Fase 1: Obtención de la información: definición del ámbito de interés y obtención de la información (en nuestro caso a través de las conversaciones de Twitter).
- Fase 2: Análisis de la información.
- Fase 3: Resumen de la información.
- Fase 4: Elaboración de los *insights*.

Tal y como señalamos en los siguientes puntos, consideramos que SETAL presenta diversas mejoras respecto a otros sistemas:

Mejor representatividad de la muestra. Las reuniones de grupo se realizan con un máximo recomendado de unos diez miembros. SETAL puede identificar a miles de usuarios. En este trabajo de investigación llegamos a identificar a 4.523 *influencers* de los cuales trabajamos finalmente con una cifra cercana a 400.

Desaparición del sesgo. SETAL se basa en el método del análisis de contenido y la técnica de la netnografía, de tal manera que el entrevistador no interviene y elimina el sesgo, analizando las conversaciones que realizan los usuarios de forma totalmente espontánea y libre en la red social Twitter.

Aplicación de filtros. SETAL es capaz de eliminar el “ruido” existente para centrar el análisis en aquellos contenidos que pueden aportar valor al estudio.

Análisis gramatical preciso. El nivel de profundidad de análisis automático supera la capacidad humana en cantidad y precisión. El algoritmo que desarrolla SETAL permite identificar aquellas palabras, oraciones y usuarios que son realmente “importantes” (por su coocurrencia con los otros contenidos analizados, las propiedades de la Teoría de grafos y su distribución estadística).

Medición de los sentimientos. SETAL incorpora un diccionario propio adaptado para detectar las palabras y estructuras que contienen la expresión de sentimientos. Como hemos apuntado

anteriormente, SETAL analiza no solo las palabras, incorpora también bigramas y tiempos gramaticales para la puntuación automática de la carga emocional de los textos analizados.

Resumen y jerarquización de la información. SETAL organiza toda la información de tal modo que permite identificar con facilidad la información más relevante, presentándola de forma ordenada de mayor a menor. El nivel de detalle es tal que permite elaborar la “huella digital” del usuario de los contenidos analizados.

Posibilidad de la obtención del *insight*. El análisis de SETAL nos permite reducir o incluso prescindir de la intuición, la iluminación o el llamado *momento de descubrimiento*. Nuestra propuesta considera una vía alternativa más científica obtenida a través del análisis, la reflexión y la deducción. Ello dependerá mismamente de las capacidades de los investigadores y publicitarios a la hora de realizar dichas tareas.

Finalmente, consideramos que este trabajo de investigación sugiere nuevas líneas de investigación que procuraremos continuar explorando en un futuro inmediato.

Por un lado, esperamos acabar el desarrollo de la herramienta SETAL para ofrecer un entorno de gestión más autónomo y ágil de cara a que las empresas puedan utilizarlo con facilidad y de manera frecuente. Buscaremos ampliar la capacidad de SETAL para obtener información de otras fuentes, como el acceso a las cuentas de los usuarios de Facebook u otras redes sociales, blogs y aquellas fuentes de acceso público vinculadas a los entornos de redes sociales digitales.

Dentro de nuestras prioridades futuras, aspiramos a profundizar en el perfeccionamiento del algoritmo que permite el proceso automático de identificación de los *influencers* así como la aplicación de los filtros requeridos para la eliminación del ruido y la obtención del contenido relevante.

Como afirmaban los profesionales entrevistados en la investigación cualitativa, la especial atención a las posibles mejoras en el análisis semántico contribuirán a aprovechar palabras y

oraciones que la riqueza del lenguaje humano nos proporciona y que para los ordenadores resulta difícil y complejo de reconocer correctamente. El humor, los dobles sentidos y la ironía bien identificados pueden proporcionarnos más claves para encontrar *insights*. Gracias a los trabajos que hemos tenido como referencia, al esfuerzo de todos los contribuyentes del ámbito académico dedicado a desarrollar los métodos analizados, a la fuerte implicación de las empresas en sus investigaciones y a la aplicación de la herramienta SETAL, hemos logrado situarnos en un tramo del camino que nos predispone con optimismo y esperanza a avanzar hacia un futuro en el que la comunicación publicitaria sigue conectada con una sociedad en constante evolución y volcada en los grandes avances tecnológicos de los últimos tiempos.

10. BIBLIOGRAFÍA

10. BIBLIOGRAFÍA

1. GRODEN, Claire. *More Americans are Getting their News on Facebook, Twitter* [en línea]. Disponible en: <<http://time.com/3957778/twitter-facebook-news-sources/>>. [Consulta: 04-2015].
2. *Leading Social Networks Worldwide as of August 2015, Ranked by Number of Active Users (in Millions)* [en línea]. Disponible en: <<http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>>. [Consulta: 04-2015].
3. DEL FRESNO, Miguel. *Netnografía*. Barcelona: UOC, 2011.
4. HAMMERSLEY, Martyn. *Ethnography: problems and prospects*. UK: The Open University, 2006.
5. NELSON, Michelle. *Exploring cross-cultural ambivalence: A netnography of intercultural wedding message boards*. IL, USA: University of Illinois, 2005.
6. KOZINETS, Robert. *Netnography: Doing ethnography research online*. London: SAGE, 2010.
7. DE VALCK, Kristine. *Virtual communities: A marketing perspective*. Paris: HEG School of Management, 2009.
8. Colaboradores de Wikipedia. *Folcsonomía* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <<https://es.wikipedia.org/wiki/Folcsonom%C3%ADa>>. [Consulta: 04-2014].
9. VAN DER WAL, Thomas. *Folksonomy* [en línea]. 2007. Disponible en: <<http://vanderwal.net/folksonomy.html>>. [Consulta: 04-2014].
10. BUSQUET, Jordi. *La recerca en comunicació. Què hem de saber?* Barcelona: UOC, 2006.
11. CEA, M.A. *Metodología cuantitativa: estrategias y técnicas de investigación social* Madrid, 1998.
12. SIEBER, S.D. *The Integration of Fieldwork and Survey Methods*: 337. Chicago: The University of Chicago Press, 1973.
13. Colaboradores de Wikipedia. *Social network analysis* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <<https://es.wikipedia.org/wiki/Folcsonom%C3%ADa>>. [Consulta: 04-2014].
14. *El Matemático Que Inventó Hace Más De 150 Años Cómo Buscar En Google* [en línea]. Disponible en: <<http://www.lanacion.com.ar/1761359-el-matematico-que-invento-hace-mas-de-150-anos-como-buscar-en-google>>. [Consulta: 04-2015].
15. PRELL, Cristina. *Social Network Analysis, History, Theory and Methodology* [en línea]. 2012: 62. Disponible en: <<http://www.socresonline.org.uk/11/4/prell.html>> [Consulta: 04-2014].
16. WELLMAN, Barry. *Network analysis: Some basic principles*. American Sociological Association, 1983.
17. PRELL, Cristina. *Social Network Analysis, History, Theory and Methodology* [en línea]. 2012: 76-118. Disponible en: <<http://www.socresonline.org.uk/11/4/prell.html>> [Consulta: 04-2014].
18. DEERWESTER, Scott et al. *Indexing by latent semantic analysis*. Chicago: University of Chicago, 1988.
19. PENNEBAKER, James, BOOTH, Roger y FRANCIS, Martha. *Linguistic Inquiry and Word Count* [en línea]. Disponible en: <<http://www.liwc.net/>>. [Consulta: 03-2014]
20. GOTTSCHALK, L.A y GLASER, G.C. *The measurement of psychological states through the content analysis of verbal behavior*. Berkeley: University of California Press, 1969.
21. ROSENBERG, S.D y TUCKER, G.J. *Verbal behavior and schizophrenia: The semantic dimension*. Archives of General Psychiatry, 1978.

22. STILES, W.B. *Describing talk: A taxonomy of verbal response modes*. California: SAGE, 1992.
23. PENNEBAKER, James. *Writing About Emotional Experiences as a Therapeutic Process*. Dallas: Southern Methodist University, 1997.
24. SMYTH, J.M. *Written emotional expression*. Journal of Consulting and Clinical Psychology, 1997
25. PENNEBAKER, James y FRANCIS, Martha. *Cognitive, emotional, and language processes in disclosure*, Dallas: Southern Methodist University, 1996.
26. PENNEBAKER, James, MAYNE, Tracy y FRANCIS, Martha. *Linguistic predictors of adaptive bereavement*. Journal of Personality and Social Psychology, Vol 72(4), 1997.
27. PENNEBAKER, James y FRANCIS, Martha. *LIWC: Linguistic inquiry and word count*. TX: Southern Methodist University, 2001.
28. SCHAPIRE, Rob. *Theoretical Machine Learning* [en línea]. 2008. Disponible en: <http://www.cs.princeton.edu/courses/archive/spr08/cos511/scribe_notes/0204.pdf>. [Consulta: 04-2014].
29. *MOA (Massive Online Analysis)* [en línea]. Disponible en: <<http://moa.cms.waikato.ac.nz/>>. [Consulta: 04-2014].
30. KOZINETS, Robert. *The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities*. Evanston: Northwestern University, 2002.
31. PANG, Bo y LEE, Lillian. *Opinion mining and sentiment analysis*. Foundations and Trends in Information Retrieval, 2008.
32. *Consulta la importancia de una página con PageRank* [en línea]. Disponible en: <<https://support.google.com/toolbar/answer/79837?hl=es>>. [Consulta: 04-2014].
33. ALLEN, R.B. *Several Studies on Natural Language and Back-propagation*. First International Conference on Neural Network. San Diego: IEEE, 1987.
34. HARRIS, M.D. *Introduction to Natural Language Processing*. Reston Publishing Company, 1985.
35. TOFFLER, Alvin. *The Third Wave: The Classic Study of Tomorrow*. New York: Bantam, 1980.
36. RITZER, George, DEAN, Paul y JURGENSON, Nathan. *The Coming of Age of the Prosumer* [en línea]. 2012. Disponible en: <<http://abs.sagepub.com/content/56/4/379.refs.html>>. [Consulta: 04-2014].
37. NEAPOLITAN, Jerry. *A Test of Simple Computer-Assisted Instructional Software*. American Sociological Association, 1989.
38. STEPHENSON, T.A. *An introduction to Bayesian Network Theory and Usage*. IDIAP, 2000.
39. TRAVERS, Jeffrey. *An Experimental Study of the Small World Problem*. American Sociological Association, 1969.
40. MELVILLE, Prem. *Social Media Analytics: Channeling the Power of the Blogosphere for Marketing Insight*. New York, 2009.
41. WASSERMAN, Stanley y FAUST, Katherine. *Social Network Analysis: Methods and Application: 4*. University of North Carolina, 1999.
42. QUALMAN, Erik. *Social Media Video* [en línea]. 2013. Disponible en: <<http://www.socialnomics.net/2013/01/01/social-media-video-2013/>>
43. ZHANG, Lei y WANQING, Tu. *Six Degrees of Separation on Online Society* [en línea]. 2009. Disponible en: <http://journal.webscience.org/147/2/webosci09_submission_49.pdf>. [Consulta: 04-2014].
44. WATTS, Duncan. *Six Degrees: The Science of a Connected Age*, 2004. Disponible en: <<http://www.amazon.com/Six-Degrees-The-Science-Connected/dp/0393325423>>. [Consulta: 04-2014].
45. ROUSE, Margaret. *Text mining (text analytics) definition* [en línea]. 2015. Disponible en: <<http://searchbusinessanalytics.techtarget.com/definition/text-mining>>

46. NENKOVA, Ani y MCKEOWN, Kathleen. *Automatic Summarization* [en línea]. 2011. Disponible en: <<http://www.cis.upenn.edu/~nenkova/1500000015-Nenkova.pdf>>. [Consulta: 04-2014].
47. CODINA, Lluís. *La web semántica*. Barcelona: UPF, 2006.
48. BERNERS-LEE, Tim et al. *The Semantic Web* [en línea]. 2001. Disponible en: <http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf>
49. BRAVO, Crescencio y REDONDO, M.A. *Sistemas interactivos y colaborativos en la web* [en línea]. 2004. Disponible en: <https://books.google.es/books?hl=es&lr=&id=N8qBAgAAQBAJ&oi=fnd&pg=PA195&dq=web+semantica&ots=jJgpR8P8l_&sig=t8NPU1qA6VxfmFOd5cRYphu3HC8#v=onepage&q=web%20sem%C3%A1ntica&f=false>. [Consulta: 04-2014].
50. Person of the year: you. *Times* [en línea]. Vol 168(26), 2006. Disponible en: <<http://content.time.com/time/covers/0,16641,20061225,00.html>>.
51. MCCRAKEN, Harry. TIME's Machine of the Year, 30 Years Later. *Times*. [en línea]. 2014. Disponible en: <<http://techland.time.com/2013/01/04/times-machine-of-the-year-30-years-later/>>.
52. *People of the Year: Bloggers*. [en línea]. 2004. Disponible en: <<http://abcnews.go.com/WNT/PersonOfWeek/story?id=372266>>.
53. WIMMER, Roger. *Mass Media Research:2*. Cengage Learning, 2013.
54. ELLISON, Nicole B., et al. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 2007.
55. TAPSCOTT, Anthony D. *Wikinomics: la nueva economía de las multitudes inteligentes*. Paidós, 2007.
56. JENKINS, Harry. *Convergence Culture: Where Old and New Media Collide*. NYU Press, 2006.
57. O'REILLY, Tim. *Design Patterns and Business Models for the Next Generation of Software*. [en línea]. 2005. Disponible en: <<http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>>.
58. *What is free software? The Free Software Definition*. [en línea]. Disponible en: <<http://www.gnu.org/philosophy/free-sw.html>>.
59. RODRÍGUEZ, Delia. Verdades y mentiras de los 'Trending topics'. *El País*. [en línea]. 2011. Disponible en: <http://elpais.com/diario/2011/05/15/domingo/1305431563_850215.html>.
60. COOPER, B.B. *The Surprising History of Twitter's Hashtag and 4 Ways to Get the Most out of Them*. [en línea]. 2013. Disponible en: <<https://blog.bufferapp.com/a-concise-history-of-twitter-hashtags-and-how-you-should-use-them-properly>>.
61. O'REILLY, Tim. *What Is Web 2.0 Design Patterns and Business Models for the Next Generation of Software*, 2006.
62. PRENTICE, Rachel. The anatomy of surgical simulations: The mutual articulation of bodies in and through the machine. *Social Studies of Science*, v.35(6), 2005.
63. STYHRE, Alexander. The cinematic mode of organizing: Media and the problem of attention in organization theory. *Information and Organization*, v.19(1), 2009.
64. KRUMM, John. User-generated content. *IEEE Pervasive Computing*, (4), 2008.
65. TOMÉ, Pepe. *Conecta!*: 46. Barcelona: Llorenç Rubió, 2011.
66. DEL FRESNO, Miguel. Comprendiendo los social media y mass media: un modelo para el estudio de la comunicación interpersonal colectiva en tiempos de Internet. *Nueva Época*, (11):101-102, 2012.
67. KAPLAN, Andreas. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, v.53(1), 2010.
68. LÓPEZ-FERNÁNDEZ, Rafael. *Diferencias entre medio social y red social*. [en línea]. 2013. Disponible en: <<http://marketingdigitaldesdecero.com/2013/03/17/diferencias-entre-medio-social-y-red-social>>.

69. ESCALONA, Nùria. #DIGITALK. *Del màrqueting directe a la comunicació 2.0.*: 55. Barcelona: UOC, 2013.
70. FREIRE, Juan. Redes sociales:¿ modelos organizativos o servicios digitales?.*El profesional de la información*, v.17(6): 585, 2008.
71. GARCÍA-UCEDA, Mariola. *Las claves de la publicidad*. 2008.
72. HAUSMAN, Angela. *16 Differences Between Social Media and Traditional Media*, 2012.
73. DOMENECH, Jaime. *Las opiniones de Amazon tienen la misma importancia que las críticas profesionales*. [en línea]. 2012. Disponible en: <<http://www.theinquirer.es/2012/05/16/las-opiniones-de-amazon-tienen-la-misma-importancia-que-las-criticas-profesionales.html>>.
74. SANTO, Carmen. *La influencia social cada vez se evidencia más en la compra de regalos*. [en línea]. 2013. Disponible en: <<http://www.puomarketing.com/76/16936/influencia-social-cada-evidencia-compra-regalos.html>>.
75. SÁNCHEZ-REVILLA, M.A. *Estudio INFOADEX de la inversión publicitaria en España*. [en línea]. 2014. Disponible en: <http://www.infoadex.es/infoadex_resumen_est_inv_2013.pdf>.
76. *Key Digital Trends*. [en línea]. 2014. Disponible en: <<http://www.emarketer.com/Webinar/Key-Digital-Trends-2014/4000075>>.
77. NELSON-FIELD, Karen. Why reaching fever fans won't (can't) diminish engagement. *Institute for marketing science*. Australia, 2012.
78. MAYER, R.E et al. *The nature of insight*. Massachusetts: Bradford Books, 1996.
79. DORON, Roland. *Diccionario Akal de psicología*: 112. Madrid: AKAL, 1998.
80. VALDIVIESO, Sergio. El insight en psicoanálisis y sus dimensiones. *Revista chilena de neuro-psiquiatría*, v.40(4), 2002.
81. Webster's Encyclopedic. *Insight*:735. Publishing, 1996.
82. BULLMORE, Jeremy. *Apples, Insights and Mad Inventors: and Entertaining Analysis of Modern Marketing*. England, 2006.
83. STEEL, Jon. *Truth, Lies and Advertising. The Art of Account Planing*:XIII. Wiley, 1998.
84. HARTMANN, Johannes. How to generate global insights to create global brands. *Consumer Insight*: 100. Milán, 2007.
85. LAUBOURNE, Peter. *Qualitative Research*:33. París: ESOMAR, 2007.
86. BALANZÓ, Cristina. Los consumer insight y las neurociencias: la relación entre las variables neurobiológicas y la eficacia publicitaria. *Universidad Ramón Llull, Facultat de Comunicació Blanquerna*, v.1(29), p.91-94, 2011.
87. NICOLÁS CARRILLO, Jerónimo. *El Insight creativo*: 146. Universidad de Murcia, 1999.
88. LÓPEZ VÁZQUEZ, Belén. Publicidad Emocional. *Estrategias creativas*: 40. Madrid: ESIC Editorial, 2007.
89. LEE, M.S. Anti-Consumption and Brand Avoidance. *Journal of Business Research*:171, v.62(2), 2008.
90. BECKLEY, Jacqueline et al. *Product innovation toolbox: a field guide to consumer understanding and research*, 2012.
91. WILLIAMS, Marsha. Insights vs. Findings. *Consumer Insights*:10. ESOMAR, 2007.
92. GARDNER, M. Aha! A two volume collection: aha! Gotcha aha. *Insight*, 2006.
93. ROMO, Manuela. *Psicología de la creatividad*: 55. Barcelona: Paidós, 2009.
94. TAYLOR, Irving. *Perspectives in Creativity*:225. Transaction Publishers, 2009.
95. CSIKSZENTMIHALYI, Mihaly. *Creatividad: el fluír y la psicología del descubrimiento y la invención*: 131. Paidós, 1998.

96. FORTINI-CAMPBELL, Lisa. *Hitting the Sweet Spot*. Copy Workshop, 2001.
97. VITALE, Dona. *Consumer Insights 2.0: How Smart Companies Apply Customer Knowledge to the Bottom Line*: 3. Paramount Market Publishing, 2006.
98. BERGANZA CONDE, M.R. *Investigar en comunicación. Guía práctica de métodos y técnicas de investigación social en Comunicación*. Madrid: McGraw-Hill, 2005.
99. MANHEIM, Jarol. *Análisis político-empírico. Métodos de investigación en ciencia política*. Madrid: Alianza, 1998.
100. ALVIRA MARTÍN, F.R. *Metodología de la evaluación de programas*. Centro de Investigaciones Sociológicas (CIS):54, 1991.
101. MONZÓN ARRIBAS, Cándido. *La opinión pública. Teoría, concepto y métodos*:164. Madrid: Tecnos, 1987.
102. STOETZEL, Jean. *Las encuestas de opinión pública*, Madrid: Instituto de Opinión Pública, 1973.
103. HAGUE, Paul. *Cómo hacer investigación de mercado*. Bilbao, 1994.
104. SIERRA BRAVO, Restituto. *Técnicas de investigación social*. Madrid: Paraninfo, 1982.
105. BERELSON, Bernard. *Content analysis in communication research*. Free Press, 1952.
106. HOLSTI, O.R. *Content analysis for the social sciences and humanities*. Addison-Wesley Pub. Co., 1969.
107. KRIPPENDORFF, Klaus. *Metodología de análisis de contenido*. Paidós, 1990.
108. BÁEZ y PÉREZ DE TUDELA, Juan. *Investigación cualitativa*: 290. ESIC, 2009.
109. BERIAIN, Ana. *Radiografía de la metodología en investigación publicitaria*. Publicidad, Facultad de comunicación Blanquerna, 2011.
110. MARTÍNEZ, J. *La investigación en marketing*:949. Barcelona: AEDEMO, 2000.
111. ALLOZA, Ángel et al. Marco general: ¿Cómo funciona la publicidad en televisión? *Martínez, J.[et al.]*. *La investigación en marketing*: 856. Barcelona: AEDEMO, 2000.
112. GONZÁLEZ, M.A. *Curso de publicidad*: 139. Madrid: Eresma & Celeste, 1998.
113. BERIAIN, Ana. Las constantes vitales de las marcas. Monitorización de la Salud de las marcas a través de los estudios de seguimiento continuo de la publicidad. *Actas IV Congreso Internacional Latina de Comunicación Social: Comunicación, control y resistencias*. Sociedad Latina de Comunicación Social, 2012.
114. GARCÍA FERRANDO, Manuel. *El análisis de la realidad social. Métodos y técnicas de investigación*. Madrid: Alianza Editorial, 1989.
115. SCHWARTZ, Howard. *Sociología cualitativa: Método para la reconstrucción de la realidad*: 21. México: Trillas, 1984.
116. DENZIN, Norman. *The SAGE Handbook or qualitative research*. Sage, 2011.
117. RUIZ OLABUÉNAGA, J.I. *Metodología de la investigación cualitativa*:122. Bilbao: Universidad de Deusto, 1999.
118. AEDEMO, *Códigos y guías ESOMAR y normas aplicables a la investigación de mercados*. Barcelona: AEDEMO, 2000.
119. FOX, Richard. Mail survey response rate, a meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly*: 468, v.52(4), 1988.
120. WEITZMAN, E.A. *Software and Qualitative Research*: 804. SAGE, 2000.
121. JOHNSON, R.R. *Estadística elemental: lo esencial*. Sage, 2008.
122. PREECE, Jennifer. *Interaction design: beyond human-computer interaction*. New York, 2015.
123. BRADBURN, Norman. *Asking questions: the definitive guide to questionnaire design--for market research, political polls, and social and health questionnaires*. John Wiley & Sons, 2004.

124. VILCHES, Lorenzo. *La investigación en comunicación: métodos y técnicas en la era digital*. GEDISA, 2011.
125. GOEBERT, Bonnie. *Beyond Listening: Learning the Secret Language of Focus Group*: 214. New York, 2002.
126. GUTIÉRREZ BRITO, Jesús. ¿Prolongación, variación o ruptura con el focus group? *Revista Electrónica de Epistemología de Ciencias Sociales*, (41), 2011.
127. KAMBERELIS, George. *Focus Groups: From Structured Interviews to Collective Conversations*:3. Routledge, 2013.
128. LIAMPUTTONG, Pranee. *Focus Group Methodology: Principle and Practice*: 42. SAGE, 2011.
129. MUCCHIELLI, Roger. *L'interview de groupe: partie Connaissance du problème, partie Applications pratiques*, 1974.
130. IBÁÑEZ, Jesús. Perspectivas de la investigación social: el diseño en las tres perspectivas. *El análisis de la realidad social. Métodos y técnicas de la investigación social*. Madrid, 2000.
131. ORTÍ, Alfonso. *La apertura y el enfoque cualitativo o estructural: la entrevista abierta y la discusión de grupo*: 198. Alianza, 1989.
132. MORGAN, D.L. *Focus Groups as qualitative research*. SAGE, 1991.
133. CALLEJO, Javier. *El grupo de discusión: introducción a una práctica de investigación*: 25. Ariel, 2001.
134. PERCY, Larry. A Theory-Based Approach to Pretexting Advertising. *William D. Wells Book: Measuring Advertising Effectiveness*. Chicago, 1997.
135. MORIARTY, S.M. *Advertising: Principles and Practice*, 2014.
136. LINSTONE, H.A. *The Delphi Method: Techniques and Applications*. Adisson-Wesley Pub. Co., 1975.
137. LANDETA RODRÍGUEZ, John. *El método Delphi: Una técnica de previsión del futuro*:3. Barcelona, 2002.
138. GORDON, T.J. *The Delphi Method Research, Futures Research Methodology*. Millenium Project, 1994.
139. DALKEY, Norman. *The Delphi Method and Experimental Study of Group Opinion*:16. RAND Corporation, 1969.
140. STRAKER, Karla et al. *Comparing and complementing methods: traditional market research vs. deep customer insights*: 5048. Tokyo, 2013.
141. ZALTMAN, Gerald. *How customers think: Essential insight into the mind of the markets*. Boston: Harvard Bussiness School Press, 2003.
142. SEN, Prit. Market Research Costs How Much? *Franchising World*, v.41(10), 2009.
143. GRIFFIN, Abbie. The Voice of the Customer. *Marketing Science*, v.12(1), 1993.
144. GRAVES, Philip. *Consumer.ology: The Market Research Myth, the Truth about Consumer and the Psychology of Shopping*:4. Nicholas Brealey, 2010.
145. WITELL, Lars. Idea generation: customer co-creation versus traditional market research techniques. *Journal of Service Management*, v.22(2), 2011.
146. GOFFIN, Keith. Beyond the voice of the customer: Ethnographic market research. *Research-Technology Management*, v.55(4), 2012.
147. TROTT, Paul. The role of market research in the development of discontinuous new products. *European Journal of Innovation Management*, v.4(3), 2001.
148. FRAM, Eugene. The shopping list studies and projective techniques: a 40-year view. *Marketing Research*, v.3(4), 1991.
149. MILLER, Cyndee. Respondents Project Let Psyches Go Crazy. *Marketing News*, v.25(1), 1991.

150. CATTERALL, Miriam. Academics, practitioners and qualitative market research. *Qualitative Market Research: An International Journal*, v.1(2), 1998.
151. ROBSON, S.L. Group discussions. *BIRN Robin J. The Handbook of International Market Research Techniques*. K.P Publishers, 2000.
152. BODDY, Clive. Projective techniques in business research: are they useful. *Proceedings of the National Business Student Research Conference*, v.12(2), 2004.
153. RICHMAN, Joseph. Jokes as a projective technique: the humor of psychiatric patients. *American journal of psychotherapy*, v.50(3), 1996.
154. LILIENFELD, Scott. The scientific status of projective techniques. *Psychological science in the public interest*, v.1(2), 2000.
155. COLMAN, Andrew. *A Dictionary of Psychology*. Oxford University Press, 2015.
156. MCDANIEL, Carl. *Contemporary Marketing Research*. West Publishing Company, 1995.
157. LIDWELL, William et al. *Universal Principles of Design: 100 Ways to Enhance Usability, Influence Perception, Increase Appeal, Make Better Design Decisions, and Teach Through Design*. Rockport Publishers, 2003.
158. SRIVATSA, Achala et al. *The Case of the Elusive Insight. Consumer Insights*. ESOMAR, 2007.
159. ROWLAND, Greg. *Changing Consumer Behaviour in Russia. Consumer Insights*. ESOMAR, 2007.
160. KANTAR MEDIA [en línea]. 2015. Disponible en: <<http://www.kantarmedia.es/>>.
161. BIGGS, N.L et al. *Graph Theory*. Oxford, 1986.
162. DEO, Narsingh. *Graph Theory with applications to engineering and computer science*. PHI Learning, 2004.
163. FOULDS, L.R. *Graph Theory and Its Applications*. New York, 1992.
164. VAN DOOREN, Paul. *Graph Theory and Its Applications*. Dublin, 2009.
165. HOPKINGS, Brian. The Truth about Königsberg. *The College Mathematics Journal*, v.35(3), 2005.
166. SCOTT, John. Social network analysis: developments, advances, and prospects. *Social network analysis and mining*, v.1(1), 1991.
167. ESTRADA, Ernesto. *The structure of complex networks: theory and applications*. Oxford University Press, 2011.
168. BALABANIAN, Norman et al. *Teoría de redes eléctricas*. Reverté, 1972.
169. FRITSCH, Gerda y FRITSCH, Rudolf. *The Four-Colour Theorem*. Springer Science & Business Media, 1998.
170. GONTHIER, Georges. Formal proof: the four-color theorem. *Notices of the AMS*, v.55(11), 2008.
171. APPEL, Kenneth y HAKEN, Wolfgang. *Every planar map is four colorable*. Bull. Amer. Math. Soc, 1976.
172. Colaboradores de Wikipedia. *Teoría de los cuatro colores* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en:<https://es.wikipedia.org/wiki/Teorema_de_los_cuatro_colores>.
173. CHARTRAND, Gary. *Introductory Graph Theory*. Dover, 1977.
174. CRILLY, A.J. Arthur Cayley: Mathematician Laureate of the Victorian Age. *Johns Hopkins University Press*, 2006.
175. SYLVESTER, James. *Chemistry and Algebra*. London: McMillan and Co, 1878.
176. ZHANG, Lei. *Knowledge graph theory and structural parsing*. University of Twente, 2002.
177. SOWA, J.F. Peirce's tutorial on existential graphs. *Semiotica*, v.2011(186), 2011.
178. FREEMAN, Linton. *El Desarrollo del Análisis de Redes Sociales: Un Estudio De Sociología De La Ciencia*. Palibrio, 2012.

179. SIMMEL, Georg. *Georg Simmel on individuality and social forms*. University of Chicago Press, 2011.
180. KARINTHY, Frigyes. *Chain-links*. Everything is the Other Way, 1929.
181. SCHRAMM, H.J. *Freight Forwarded's Intermediary Role in Multimodal Transport Chains: a Social Network Approach*. Courier Corporation, 2012.
182. RADCLIFFE-BROWN, A.R. On social structure. *Journal of the Anthropological Institute of Great Britain and Ireland*, 1940.
183. GRIMALDI, Ralph. *Matemáticas discreta y combinatoria: introducción y aplicaciones*. Pearson Educación, 1998.
184. TURNER, Lynn y WEST, Richard. *Handbook of sociological theory*. Sage Publications, 2006.
185. De NOOY, Wouter et al. *Exploratory social network analysis with Pajek*. Cambridge University Press, 2011.
186. SOLOMONOFF, Ray et al. Connectivity of random nets. *The bulletin of mathematical biophysics*, v.13(2), 1951.
187. FREEMAN, Linton. La centralidad en las redes sociales. Clarificación conceptual. *Politica y sociedad*. Madrid, 2000.
188. POOL, Ithiel de Sola et al. Contacts and Influence. *Social Networks: 1*, 1978.
189. NADEL, Frederick. *SF Nadel, The Theory of Social Structure*. London: Cohen & West, 1957.
190. FREEMAN, Linton. The development of social network analysis. *A Study in the Sociology of Science*, 2004.
191. MANNING, Christopher. *Foundations of statistical natural language processing*. Cambridge, 2000.
192. BOLLOBÁS, Béla. *Modern graph theory*. Springer Science & Business Media, 1998.
193. BEAUCHAMP, Murray. *An improved index of centrality*. Behavioral Science, 1965.
194. Colaboradores de Wikipedia. *Seis grados de separación* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <https://es.wikipedia.org/wiki/Seis_grados_de_separaci%C3%B3n>.
195. WASSERMAN, Stanley. Network science: an introduction to recent statistical approaches. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009.
196. HARARY, Frank et al. *Graph Theory*. Addison-Wesley Publishing Company, 1969.
197. BERGS, Alexander. *Social networks and historical sociolinguistics: studies in morphosyntactic variation in the Paston letters (1421-1503)*. Walter de Gruyter, 2005.
198. WOODS, W.A. Transition network grammars for natural language analysis. *Communications of the ACM*, v.13(10), 1970.
199. BONACICH, Phillip. Power and centrality: a family of measures. *American Journal of Sociology*, v.92(5), 1987.
200. RABINER, L.R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, v.77(2), 1989.
201. CHARNIAK, Eugene. A Bayesian model of plan recognition. *Artificial Intelligence*, v.64(1), 1993.
202. RATNAPARKHI, Adwait. Learning to parse natural language with maximum entropy models. *Machine learning*, v.34(1-3), 1999.
203. LITMAN, Jessica. *Revising copyright law for the information age*. Or. L. Rev., 1996.
204. ROBERTS, Don D. *The existential graphs of Charles S. Peirce*. Walter de Gruyter, 1973.
205. ROBINS, G.L. New specifications for exponential random graph models. *Sociological methodology*, v.36(1), 2006.
206. BURT, Ronald. Structural holes and good ideas1. *American journal of sociology*, v.110(2), 2004.

207. WASSERMAN, Stanley. *Social network analysis: Methods and applications*. Cambridge university press, 1994.
208. WATTS, Duncan. *Collective dynamics of 'small-world' networks*, 1998.
209. Colaboradores de Wikipedia. *Modelo Watts y Strogatz* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <https://es.wikipedia.org/wiki/Modelo_Watts_y_Strogatz>.
210. BARABÁSI, Albert-Laszlo. Statistical mechanics of complex networks. *Reviews of Modern Physics*, v.74(1), 2002.
211. HOPPE, Bruce et al. Social network analysis and the evaluation of leadership networks. *The Leadership Quarterly*, v.21(4), 2010.
212. MOODY, James et al. Structural cohesion and embeddedness: A hierarchical concept of social groups. *American Sociological Review*, 2003,
213. JIN, Emily et al. Structure of growing social networks. *Physical review*, v.64(4), 2001.
214. SCHNETTLER, Sebastian. A structured overview of 50 years of small-world research. *Social Networks*, v.31(3), 2009.
215. BACKSTROM, Lars et al. *Four Degrees of Separation*. ArXiv, 2011.
216. SINGH, S. *Graph Theory*. New Delhi, 2010.
217. BIGGS, N.L. *Graph theory, 1736-1936*. Oxford, 1986.
218. BANG-JENSEN, Jørgen. *Digraphs: theory, algorithms and applications*. Springer Science & Business Media, 2008.
219. EASLEY, David. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.
220. Colaboradores de Wikipedia. *Cycle (graph theory)* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <[https://en.wikipedia.org/wiki/Cycle_\(graph_theory\)](https://en.wikipedia.org/wiki/Cycle_(graph_theory))>.
221. CALDWELL, Chris. *Graph Theory Glossary*. [en línea]. 1995. Disponible en: <<http://primes.utm.edu/graph/glossary.html>>
222. BOLLOBÁS, Béla. *Extremal graph theory*. New York, 1978.
223. Colaboradores de Wikipedia. *Centrality* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <<https://en.wikipedia.org/wiki/Centrality>>.
224. BRANDES, Ulrik. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, v.30(2), 2008.
225. Colaboradores de Wikipedia. *Betweenness centrality* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <https://en.wikipedia.org/wiki/Betweenness_centrality>.
226. BAVELAS, Alex. Communication patterns in task-oriented groups. *Journal of the acoustical society of America*, 1950.
227. FREEMAN, Linton. *A set of measures of centrality based on betweenness*. Sociometry, 1977.
228. Colaboradores de Wikipedia. *Closeness centrality* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <https://en.wikipedia.org/wiki/Centrality#Closeness_centrality>.
229. RUHNAU, Britta. Eigenvector-centrality—a node-centrality?. *Social networks*, v.22(4), 2000.
230. HAVELIWALA, T.H. The Second Heigenvalue of the Google Matrix. USA, 2003.
231. Colaboradores de Wikipedia. *Eigenvector centrality* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <https://en.wikipedia.org/wiki/Centrality#Eigenvector_centrality>.
232. OTTE, Evelien. Social network analysis: a powerful strategy, also for the information sciences. *Journal of information Science*, v.28(6), 2002.
233. SHIRINIVAS, S.G et al. Applications of graph theory in computer science an overview. *International Journal of Engineering Science and Technology*, v.2(9), 2010.

234. NEWMAN, Mark. *Networks: an introduction*. Oxford University Press, 2010.
235. BURCH, Hal y CHESWICK, Bill. *The Internet Mapping Project*. [en línea]. Disponible en: <<http://cheswick.com/ches/map/>>
236. PITTS, F.R. A graph theoretic approach to historical geography. *The Professional Geographer*, v.17(5), 1965.
237. AMARAL, L.A. Classes of small-world networks. *Proceedings of the national academy of sciences*, v.97(21), 2000.
238. SEN, D.P et al. Small-world properties of the Indian Railway network. *Physical Review*, v.67(3), 2003.
239. GASTNER, M.T. Diffusion-based method for producing density-equalizing maps. *Proceedings of the National Academy of Sciences of the United States of America*, v.101(20), 2004.
240. MAPS. *Rutas aéreas* [en línea]. 2015. Disponible en: <http://s191.photobucket.com/user/keesje_pics/media/NWApacific.jpg.html?t=1226616009>.
241. METRORAIL [en línea]. 2014. Disponible en: <<http://www.murraycavanaugh.com/~murrayca/images/metro-map-9-14.png>>.
242. GARMIN [en línea]. 2014. Disponible en: <<https://buy.garmin.com/es-ES/ES/conduciendo/descatalogados/nulink-1695-live/prod70853.html>>.
243. MILROY, James y MILROY, Lesley. Linguistic change, social network and speaker innovation. *Journal of linguistics*, v.21(2), 1985.
244. TURING, Alan. *Computing machinery and intelligence*. Mind, 1950.
245. ZHONG, Shi. Generative model-based document clustering: a comparative study. *Knowledge and Information Systems*, v.8(3), 2005.
246. ZHA, Hongyuan. *Generic summarization and keyphrase extraction using mutual reinforcement principle and sentence clustering*, ACM, 2002.
247. AGIRRE, Eneko. *Word sense disambiguation: Algorithms and applications*. Springer Science & Business Media, 2007.
248. MATSUO, Yutaka. Graph-based word clustering using a web search engine. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2006.
249. PORTES, Alejandro. *Social capital: Its origins and applications in modern sociology*. Boston, 2000.
250. KATZ, Leo. A new status index derived from sociometric analysis. *Psychometrika*, v.18(1), 1953.
251. RUSINOWSKA, Agnieszka et al. Social Networks: Prestige, Centrality and Influence. *12th International Conference in Relational and Algebraic Methods in Computer Science*, v.72(1). Rotterdam, 2011.
252. BAKSHY, Eytan, et al. The role of social networks in information diffusion. *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012.
253. STEVENSON, Mark et al. *Mining Rules for Word Sense Disambiguation*. San Leopoldo, 2005.
254. SAHA, S.K et al. *Word Clustering and Word Selection Based Feature Reduction for MaxEnt Based Hindi NER*. ACL, 2008.
255. GRIFF'S GRAPHS. *The Graph of a Social Network*. [en línea]. 2014. Disponible en: <<https://griffgraphs.wordpress.com/tag/social-network/>>.
256. PINHEIRO, André. *Social Network Analysis in Telecommunications*, 2011.
257. STATTNER, Erick. Social network analysis in epidemiology: Current trends and perspectives. *Research Challenges in Information Science (RCIS)*, 2011.
258. MURRAY, Nancy. Profiling in the Age of Total Information Awareness. *Race and Class*, v.52(2), 2010.

259. FOWLER, James H., et al. *Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study*. *Bmj*, 2008.
260. BOHN, Shawn et al. *An approach to incorporate Text into a social network analysis of communication graphs*. IEEE, 2009.
261. BRIN, Sergey et al. *The Anatomy of a Large-scale Hypertextual Web Search Engine*, 1998.
262. GOOGLE [en línea]. *Social Network*. Disponible en: <https://www.google.es/search?q=social+network&oq=social+network&aqs=chrome..69i57j0l5.2672j0j7&sourceid=chrome&es_sm=93&ie=UTF-8>. [Consulta: 19-09-2014 13:42].
263. ZHU, X.G. Getting connected: analysis and principles of biological networks. *Genes & Development*, v.21(9), 2007.
264. ANDRE, McKenzie. Transmission Network Analysis to Complement Routine Tuberculosis Contact Investigations. *AMJ Public Health*, v.97(3), 2007.
265. SPORNS, Olaf. *Small-world inside big brains*. PNAS, 2006.
266. STAM, C.J. Graph Theoretical Analysis of Complex Networks in the Brain. *Nonlinear Biomedical Physics*, v.1(3), 2007.
267. NEWMAN, P.S. Summarizing archived discussions: a beginning. *Proceedings of the 8th international conference on intelligent user interfaces*. Miami, 2003.
268. MASON, O.V et al. Graph Theory and Networks in Biology. *Systems biology*. v.1(2), 2007.
269. BUNN, A.G et al. Landscape connectivity: a conservation application of graph theory. *Journal of environmental management*, v.59(4), 2000.
270. FEATHER, John. *The Information Society: A Study of Continuity and Change*. London, 2008.
271. SWASH, Gill. *UK Bussiness Information of the Internet*. New Library Word, 1998.
272. EDMUNDS, Angela y MORRIS, Anne. The Problem of Information Overload in Bussiness Organisations: A Review of the Literature. *International Journal of Information Management*, v.20(2000), 2000.
273. LEWIS, David. *Information Overload*. London, 1999.
274. SNIJDERS, Chris et al. ““Big Data”: Big Gaps of Knowledge in the Field of Internet Science.” *International Journal of Internet Science*, v.7(1), p.1-5, 2012.
275. CARENINI, Giuseppe et al. *Methods for Mining and Summarizing Text Conversations*. University of Briths Columbia, 2011.
276. CORNELLA, Alfons. *Cómo sobrevivir a la infoxicación*. Infonomia.com, 2000.
277. SPÄRCK JONES, Karen. Automatic summarizing: factors and directions. *MIT Press, 1998*. Cambridge MA, Ed. I. Mani and M. Maybury, p.21, 1998.
278. HOVY, Eduard. *The Oxford Handbook of Computational Linguistics*, 2005.
279. DAS, Dipanjan y MARTINS, André. *A Survey on Automatic Text Summarization*, 2007.
280. TORRES-MORENO, Juan Manuel. *Automatic Text Summarization*, 2014.
281. LUHN, H. P. The Automatic Creation of Literature Abstracts. *IBM Journal of Research Development*, p.159-165, 1958.
282. Lin, C.Y. y HOVY, Eduard. Identifying Topics by Position. *Proceedings of the fifth conference on Applied Natural Language*, p.283-290. Washington, 1997.
283. EDMUNDSON, H. P. (1969). New Methods in Automatic Extracting. *Journal of the Association for Computing Machinery*, v.16(2), 1969.
284. SALTON, Gerard et al. A vector space model for automatic indexing. *Communications of the ACM*, v.18(11),1975.

285. SALTON, Gerard. *Automatic Text Processing: the transformation, analysis, and retrieval of information by computer*. Boston: Addison-Wesley Longman Publishing Co, 1988.
286. KUPIEC, J.P et al. A trainable document summarizer. *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval.*, 1995.
287. RADEV, D.R et al. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. *Proceedings of the 2000 NAACL-ANLP Workshop on Automatic Summarization*, 2000.
288. LESKOVEC, J.M et al. *Impact of linguistic analysis on the semantic graph coverage and learning of document extracts*, 2005.
289. BERGER, Adam et al. OCELOT: a system for summarizing Web pages. *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, p.144-151. Atenas, 2000.
290. BARZILAY, R.L. Modelling local coherence: An entity-based approach. *Computational Linguistics*, v.34(1), 2008.
291. BRANDOW, R.M et al. Automatic condensation of electronic publications by sentence selection. *Information Processing & Management*, v.31(5), 1995.
292. MARCU, Daniel. *The rhetorical parsing, summarization, and generation of natural language texts*, University of Toronto, 1997.
293. POLANYI, L.C et al. A rule-based approach to discourse parsing. *Proceedings of the fifth SIGdial workshop on discourse and dialogue*, 2004.
294. BARZILAY, R.L. Using lexical chains for text summarization. *Advances in automatic text summarization*. Massachusetts, 1999.
295. FUENTES, María. *A Flexible Multitask Summarizer for Documents from Diferent Media, Domain, and Language*. Barcelona: UPC, 2008.
296. MANI, Inderjeet. *Advances in Automatic Text Summarization*. Massachusetts, 1999.
297. STEINBERGER, Joseph y JEŽEK, Karel. Evaluation measures for text summarization. *Computing and Informatics*, v.29(2009), 2009.
298. TORRES, J.M. *Automatic Text Summarization*, 2014.
299. MANI, Inderjeet. *Automatic Summarization*. Amsterdam, 2001.
300. HAHN, U.M. The challenges of automatic summarization. *Computer*, v.33(11), 2000.
301. BORKO, Harold y BERNIER, Charles. *Abstracting Concepts and Methods*, 1975.
302. JEŽEK, K.S. Automatic text summarization, the state of the art 2007 and new challenges. *Proceedings of the 7th Znalosti Conference*. Bratislava, 2008.
303. SIZOV, Gleb. *Extraction-Based Automatic Summarization. Theoretical and Empirical Investigation of Summarization Techniques*. Norwegian University of Science and Technology, 2010.
304. GUPTA, V.L. A Survey of Text Summarization Extractive Techniques. *Journal of Emerging Technologies in Web Intelligence*, v.2(3), 2010.
305. HOVY, E.L. Automated text summarisation in SUMMARIST. Cambridge, 1999.
306. DANG, H.O. Overview of the TAC 2008 Update Summarization Task. *Proceedings of Text Analysis Conference*, 2008.
307. NENKOVA, A.M. Automatic Summarization. *Foundations and Trends in Information Retrieval*, v.5(2-3), 2011.
308. DANG, H.T. TAC 2010 guided summarization task guidelines. *Proceedings of Text Analysis Conference*, 2008.
309. McKEOWN, K.K et al. *Towards Multidocument Summarization by Reformulation: Progress and Prospects*. Orlando, 1999.

310. EVANS, D.K. *Identifying Similarity in Text: Multi-Lingual Analysis for Summarization*. Columbia University, 2005.
311. Hovy, E.L y LIN, C.Y. Automated text summarisation in SUMMARIST, p.81-94. Cambridge, 1999.
312. BOGURAEV, B.K. Saliency-based content characterisation of text documents. *Advances in automatic text summarization*, p.99-110, 1999.
313. ONO, K.S et al. Abstract Generation Based on Rhetorical Structure Extraction. *Proceedings of the 15th International Conference on Computational Linguistics*. Kyoto, 1994.
314. KHAN, A.U. MRST: A New Technique For Information Summarization. *The Second World Enformatika Conference*. Çanakkale, 2005.
315. KLEINBERG, J.K et al. The web as a graph: measurements, models, and methods. *Computing and combinatorics*, p.1-17, 1999.
316. BRIN, Sergey et al. *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, 1999.
317. RAE. *Anáfora* [en línea]. 2015. Disponible en: <<http://www.rae.es/diccionario-panhispanico-dedudas/terminos-linguisticos>>. [consulta: 15-07-2015].
318. Colaboradores de Wikipedia. *Hiperónimo* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <<https://es.wikipedia.org/wiki/Hiper%C3%B3nimo>>.
319. Colaboradores de Wikipedia. *Holonimia* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <<https://es.wikipedia.org/wiki/Holonimia>>.
320. KYOOMARSI, F. K et al. Optimizing text summarization based on fuzzy logic. *Seventh IEEE/ACIS International Conference on Computer and Information Science*, 2008.
321. JONES, K.S et al. *Evaluating Natural Language Processing Systems: and Analysis and Review*. Springer, 1996.
322. MINEL, J.L et al. How to Appreciate the Quality of Automatic Text Summarization. *Workshop on Intelligent Scallable Text Summarization*, New Jersey, 1997.
323. MOLLÁ, D.H. Intrinsic versus extrinsic evaluations of parsing systems. *Proceedings of EACL03 workshop on Evaluation Initiatives in Natural Language Processing*. Budapest, 2003.
324. SWETS, J.A. *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers*, 1996.
325. Colaboradores de Wikipedia. *Curva ROC* [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <https://es.wikipedia.org/wiki/Curva_ROC>.
326. DING, C.H. A Probabilistic Model for Latent Semantic Indexing. *Journal of the American Society for Information Science and Technology*, v.56(6), 2005.
327. NENKOVA, A.P. Evaluating Content Selection in Summarization: The Pyramid Method. *Document Understanding Conference*. Vancouver, 2004.
328. LANDAUER, T.K et al. An Introduction to Latent Semantic Analysis. *Discourse Processes*, v.25(2-3), 1998.
329. GONG, Yihong et al. Generic text summarization using relevance measure and latent semantic analysis. *SIGIR '01 Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, 2002.
330. MORRIS, A.H et al. The Effects and Limitations of Automated Text Condensing on Reading Comprehension Performance. *Advances in Automatic Text Summarization*. MIT Press, 1999.
331. PANG, Bo. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, p.36, 2008.
332. CUI, H.M et al. *Comparative experiments on sentiment classification for online product reviews*. Boston, 2006.

333. *E-Commerce Reviews: The dramatic impact of online reviews on consumers' purchasing journey* [en línea]. 2015. Disponible en: <<http://influence-central.com/e-commerce-reviews-the-dramatic-impact-of-online-reviews-on-consumers-purchasing-journey/>>.
334. KIM, S.M y HOVY, E.L. Determining the sentiment of opinions. *Natural Language Processing (IJCNLP)*. Jeju Island, 2004.
335. LIU, Bing. *Sentiment Analysis and Opinion Mining*. Toronto, 2012.
336. TABOADA, M.B et al. *Lexicon-Based Methods for Sentiment Analysis*. MIT Press, 2011.
337. ALAMI, N.L et al. Automatic texts summarization: current state of the art. *Journal of Asian Scientific Research*, v.5(1), 2015.
338. NEWSBLASTER. *Columbia Newsblaster*. [en línea]. 2015. Disponible en: <<http://newsblaster.cs.columbia.edu/>>.
339. HU, Meishan. Comments-oriented blog summarization by sentence extraction. *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. Lisboa, 2007.
340. CHAKRABARTI, D.P. Event Summarization using Tweets. *Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM)*. Barcelona, 2011.
341. SUN, J.T et al. Web-page summarization using clickthrough data. *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, p.194-201, 2005.
342. BUYUKKOKTEN, O.G et al. Text Summarization of Web pages on Handheld Devices. *Proceedings of the workshop on Automatic Summarization*. Pittsburgs, 2001.
343. Colaboradores de Wikipedia. RSS [en línea]. *Wikipedia, la enciclopedia libre*. Disponible en: <<https://es.wikipedia.org/wiki/RSS>>.
344. GOOGLE. *Resumen automático de textos*. [en línea]. 2015. Disponible en: <https://www.google.es/search?q=resumen+autom%C3%A1tico+de+textos&oq=resumen+autom%C3%A1tico+de+textos&aqs=chrome..69i57j0l2.5030j0j9&sourceid=chrome&es_sm=93&ie=UTF-8>. [consulta: 13-11-2014 16:08].
345. RAMBOW, Owen et al. Summarizing Email Threads. *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*. Boston, 2004.
346. BIADSY, F.H. *An Unsupervised Approach to Biography Production Using Wikipedia*. ACL, 2008.
347. BANKO, M.M. Headline generation based on statistical translation. *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*. AFC, 2000.
348. YOUSFI-MONOD, M.F et al. Supervised machine learning for summarizing legal documents. *Advances in Artificial Intelligence*, p.51-62, 2010.
349. LIU, Ling y ÖZSU, Tamer. *Encyclopedia of Database Systems. Opinion Mining*, 2009.
350. MURESAN, Smaranda et al. Combining Linguistic and Machine Learning Techniques for Email Summarization. *Proceedings of the 2001 Workshop on Computational Natural Language Learning*, v.7(19), 2001.
351. WAN, Stephen y McKeown, K.K. Generating overview summaries of ongoing email thread discussions. *Proceedings of the 20th International Conference on Computational Linguistics*. AFC, 2004.
352. GARERA, Nikesh y YAROUSKY, David. Modeling latent biographic attributes in conversational genres. *47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. Suntec, 2009.
353. CARENINI, Giuseppe et al. Summarizing Emails with Conversational Cohesion and Subjectivity. *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Ohio, 2007.

354. GUPTA, Alok et al. An Empirical Study of Consumer Switching from Traditional to Electronic Channels: A Purchase-Decision Process Perspective. *International Journal of Electronic Commerce*, v.8(3), 2004.
355. KU, Lun-Wei et al. Opinion Extraction, Summarization and Tracking in News and Blog Corpora. *AAAI spring symposium: Computational approaches to analyzing weblogs*. AAAI, 2006.
356. KOLCZ, Aleksander et al. Summarization as feature selection for text categorization. *Proceedings of the tenth international conference on Information and knowledge management*. Atlanta, 2001.
357. ASBAGH, M.J et al. *Blog Summarization for Blog Mining*, 2009.
358. CARBONELL, Jaime y GOLDSTEIN, Jade. The use of MMR, diversity-based reranking for reordering documents and producing summaries. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. Melbourne, 1998.
359. ZHOU, Lei y HOVY, E.H. On the Summarization of Dynamically Introduced Information: Online Discussions and Blogs. *Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs*. Stanford, 2006.
360. SAYYADI, H.H et al. Event Detection and Tracking in Social Streams. *Proceedings of the Third International ICWSM Conference*, p.311-314, 2009.
361. POPESCU, A.M y PENNACCHIOTTI, Marco. Detecting Controversial Events from Twitter. *Proceedings of the 19th ACM international conference on Information and knowledge management*. Toronto, 2010.
362. WATANABE, K.O et al. Jasmine: a real-time local-event detection system based on geolocation information propagated to microblogs. *Proceedings of the 20th ACM international conference on Information and knowledge management*. Glasgow, 2011.
363. SAKAKI, T.O et al. Earthquake shakes Twitter users: real-time event detection by social sensors. *Proceedings of the 19th international conference on World wide web*. North Caroline, 2010.
364. CHUA, F.C et al. Automatic Summarization of Events from Social Media. *International AAAI Conference on Web and Social Media*, 2013.
365. BEAUMONT, Claudine. *New York plane crash: Twitter breaks the news, again*. [en línea]. 2009. Disponible en: <<http://www.telegraph.co.uk/technology/twitter/4269765/New-York-plane-crash-Twitter-breaks-the-news-again.html>>.
366. TELEGRAPH, T. *New York plane crash: Twitter breaks the news, again*. [en línea]. 2014. Disponible en: <<http://www.telegraph.co.uk/technology/twitter/4269765/New-York-plane-crash-Twitter-breaks-the-news-again.html>>.
367. CHAKRABARTI, Deepayan y PUNERA, Kunal. Event Summarization Using Tweets. *Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM)*. Barcelona, 2011.
368. SHARIFI, Beaux et al. Summarizing Microblogs Automatically. *Experiments in Microblog Summarization*. Colorado, 2010.
369. ANSI, *American National Standard for Writing Abstracts*. New York, 1979.
370. GOLEMAN, Daniel. *Inteligencia emocional*. Barcelona: Editorial Kairós, 1996.
371. MUSTO, Cataldo et al. A comparison of Lexicon-based approaches for Sentiment Analysis of microblog posts. *Information Filtering and Retrieval*, v.59, 2014.
372. LIU, Bing. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. New York: 2015.
373. DAS, S.R y CHEN, M.W. Yahoo! for Amazon: Sentiment Parsing from Small Talk on the Web. *EFA 2001 Barcelona Meetings*. Barcelona: Social Science Research Network, 2001.
374. TONG, R.M. An operational system for detecting and tracking opinions in on-line discussions. *Working Notes of the SIGIR Workshop on Operational Text Classification (OTC)*. New Orleans, 2001.
375. PANG, Bo et al. Thumbs up? Sentiment Classification using Machine Learning Techniques. *Empirical methods in natural language processing*. Philadelphia, 2002.

376. TURNEY, P.D. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Philadelphia, 2002.
377. NASUKAWA, Tetsuya et al. Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques. *Proceedings of the Third IEEE International Conference on Data Mining*. I.C Society, 2003.
378. YU, Hong y HATZIVASSILOGLU, Vasileios. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. *Proceedings of the 2003 conference on Empirical methods in natural language processing*. Sapporo, 2003.
379. DAVE, Kushal, LAWRENCE, Steve y PENNOCK, David. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. *Proceedings of the 12th international conference on World Wide Web*. Budapest, 2003.
380. LIU, Bing. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*, 2007.
381. PANG, Boo y LEE, Lillian. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Michigan, 2005.
382. KAHNEMAN, Daniel. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.
383. DE SOUSA, Ronald. *Emotion*. [en línea]. 2013. Disponible en: <<http://plato.stanford.edu/entries/emotion/>>. [consulta: 14-08-2015]
384. JAMES, William. What is an emotion? *Oxford University Press. Mind Association*, v.9(34), 1884.
385. RYLE, Gilbert. *The Concept of Mind*. Chicago, 1949.
386. RUSSELL, J.A y BARRET, Lisa. Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant. *Personality and Social Psychology*, v.76(5), 1999.
387. AVERILL, J.R. *A constructivist view of emotion*. London: Academic Press, 1980.
388. HARRE, Rom. *The Social Construction of Emotions*. Blackwell Pub, 1989.
389. IZARD, C.E. *The face of emotion*. New York, 1971.
390. DAMASIO, Antonio. *Y el cerebro creó al hombre*. Barcelona: Destino, 2010.
391. DELGADO LINARES, Inmaculada. *Destrezas sociales*. Ediciones Paraninfo, 2010.
392. SCHERER, K.R. What are emotions? And how can they be measured? *Social Science Information SSI*, v.44(4), 2005.
393. EKMAN, Paul. *Basic Emotions*. Chichester, 1999.
394. PLUTCHIK, Robert. *The Emotions*. University Press of America, 1991.
395. PLUTCHIK, Robert. The Nature of Emotions Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, v.89(4), 2001.
396. PARROTT, G.W. *Emotions in Social Psychology: Essential Readings*, 2001.
397. COWIE, Roddie y CORNELIUS, R.R. Describing the emotional states that are expressed in speech. *Speech Communication*, v.40(1-2), 2003.
398. VANDEBOS, G.R. *APA Dictionary of Psychology*. Washington DC: APA, 2007.
399. FRANCISCO, Virginia. *Identificación automática del contenido afectivo de un texto y su papel en la presentación de información*. Madrid: Universidad Complutense de Madrid, 2008.
400. GONÇALVEZ, P.B et al. *Panas-t: A psychometric scale for measuring sentiments on twitter*. ArXiv preprint, 2013.
401. MAYNARD, Diana y FUNK, Adam. Automatic detection of political opinions in tweets. *The semantic web: ESWC 2011 workshops*. Heraklion, 2011.

402. HEMALATHA, I.V et al. *The Second Eigenvalue of the Google Matrix*. USA, 2003.
403. PAK, Alexander et al. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *LREC*, v.10, 2010.
404. HU, Xia et al. Exploiting social relations for sentiment analysis in microblogging. *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 2013.
405. SHALUNTS, G.B et al. Sentiment Analysis of German Social Media Data for Natural Disasters. *Proceedings of the 11th International ISCRAM Conference*. USA, 2014.
406. MEDHAT, W.H et al. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, v.5(4), 2014.
407. JOACHIMS, Thorsten. A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization. *CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER SCIENCE*, 1996.
408. CHEN, C.C et al. Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, v.50(4), 2011.
409. LI, Y.M y LI. T.Y. Deriving market intelligence from microblogs. *Decision Support Systems*, v.55(1), 2013.
410. OSGOOD, Charles et al. *The Measurement of Meaning*, 1957.
411. MILLER, G.A. WordNet: a lexical database for English. *Communications of the ACM*, v.38(11), 1995.
412. FELLBAUM, Christiane. *WordNet. A lexical database for English*. [en línea]. 2015. Disponible en: <<http://wordnet.princeton.edu/>>.
413. BACCIANELLA, S.E et al. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *LREC*, v.10, 2010.
414. STRAPPARAVA, Carlo et al. WordNet Affect: an Affective Extension of WordNet. *LREC* v.4, 2004.
415. WIEBE, Janyce, WILSON, Teresa y CARDIE, Claire. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, v.39(2-3), 2005.
416. CAMBRIA, Erik y WHITE, Bruce. Jumping NLP Curves: A Review of Natural Language Processing Research. *Computational Intelligence Magazine*, v.9, 2014.
417. CAMBRIA, Erik et al. The Hourglass of Emotions. *Cognitive Behavioural Systems*. Dresden, 2012.
418. MISHNE, Gilad. Predicting Movie Sales from Blogger Sentiment. *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, 2006.
419. BERMINGHAM, Adam. On using Twitter to monitor political sentiment and predict election results. *Sentiment Analysis and Real-time Microblog Search*. Dublin, 2011.
420. DIAKOPOULOS, N.A et al. Characterizing debate performance via aggregated twitter sentiment. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010.
421. SANG, E.T.K y BOS, Johan. Predicting the 2011 Dutch Senate Election Results with Twitter. *Proceedings of SASN 2012, the EACL 2012 Workshop on Semantic Analysis in Social Networks*. Avignon, 2012.
422. ZHANG. X.F et al. Predicting Stock Market Indicators Through Twitter “I hope it is not as bad as I fear. *Procedia-Social and Behavioral Sciences*, 2010.
423. SI, J.M et al. Exploiting Topic based Twitter Sentiment for Stock Prediction. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Sofia, 2013.
424. MCGLOHON, M.G et al. Star Quality: Aggregating Reviews to Rank Products and Merchants. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 2010.
425. HONG, Yancheng y SKIENA, Steven. The Wisdom of Bookies? Sentiment Analysis Versus the NFL Point Spread. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 2010.

426. MILLER, M.S et al. Sentiment Flow Through Hyperlink Networks. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. Barcelona, 2011.
427. MOHAMMAD, Saif y YANG, Tony. Tracking sentiment in mail: How genders differ on emotional axes. *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. Portland, 2011.
428. SINCLAIR, John. *Corpus, Concordance, Collocation*, 1991.
429. MARKOV, Andrey. Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain. *Dynamic Probabilistic Systems*, 1971.
430. TWITTER. *Tweets send per day*. [en línea]. 2015. Disponible en: <<https://about.twitter.com/es/company>>.
431. TAKASHI, Tomokiyo y HURST, M. A language modeling approach to keyphrase extraction. *Proceedings of the ACL 2003 workshop on Multiword expressions*, 2003.
432. YALE, L.G et al. Trends in advertising: a look at the content of marketing-oriented journals from 1976 to 1985. *Communications of the ACM*, v.13(10), 1998.
433. LANGER, Roy. y BECKMAN, Suzanne. Sensitive research topics: netnography revisited. *Qualitative Market Research: An International Journal*, v.8(2), 2005.
434. HINE, Christine. *Virtual Ethnography*. SAGE, 2000.
435. MARKHAM, A.N. *The methods, politics, and ethics of representation in online ethnography*. Thousand Oaks, 2005.
436. QUINTON, Sara y HARRIDGE-MARCH, Sally. Relationships in online communities: The potential for marketers. *Journal of Research in Interactive Marketing*, v.4(1), 2010.
437. PACCAGNELLA, Luciano. Getting the Seats of Your Pants Dirty: Strategies for Ethnographic Research on Virtual Communities. *Journal of Computer-Mediated Communication*, v.3(1), 1997.
438. PENNEBAKER, J.W. Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, 2003.
439. FAST, L.A y FUNDER, D.C. Personality as Manifest in Word Use: Correlations with Self-report, Acquaintance report, and Behavior. *Journal of Personality and Social Psychology*, v.94(2), 2008.
440. PENNEBAKER, J.W. *The Secret Life of Pronouns: What Our Words Say About Us*. USA, 2011.
441. GIMÉNEZ DE LA PEÑA, Almudena. *Manual de prácticas de Psicología del Lenguaje*. Madrid, 2003.
442. I.A.B SPAIN. *V Estudio anual de Redes Sociales*. [en línea]. 2015. Disponible en: <<http://www.iabspain.net/wp-content/uploads/downloads/2014/04/V-Estudio-Anual-de-Redes-Sociales-versi%C3%B3n-reducida.pdf>>
443. AHKTER, J.K y SORIA, Steven. Sentiment analysis: Facebook status messages. *Technical report*. Stanford University, 2010.
444. CHA, M.H et al. Measuring User Influence in Twitter: The Million Follower Fallacy. *Fourth International AAAI Conference on Weblogs and Social Media*, 2010.
445. CATALÁN, Ignacio. *Radiografía de los principales bancos españoles*. [en línea]. 2015. Disponible en: <http://elpais.com/elpais/2015/07/03/media/1435944057_924482.html>. [consulta: 12-07-2015].

ANEXOS

ANEXOS

Diccionario LIWC

abandonado -2
deplorables -2
deplorable -2
empeorando -1
adecuada 1
abandonados -2
abandonar -2
abatido -3
abatido -2
abatidos -3
abominable -3
abominables -3
aborrece -3
aborrecemos -3
aborrecen -3
aborrecer -3
aborrecido -3
aborrecidos -3
abrazar 1
abrazo 2
abrazos 2
absolver 2
absolver 2
absolviendo 2
absolviendo 2
absorbido 1
absorbidos 1
absorto 1
absuelto 2
absuelto 2
absuelve 2
absuelve 2
absuelven 2
aburrido -3
aburrido -2
abusado -3

abusivo -3
abuso -3
abusos -3
acaparador -2
accidental -2
accidentales -2
accidentalmente -2

accidente -2
accidentes -2
acciones 1
acepta 1
aceptado -1
aceptado 1
aceptados 1
aceptar 1
aclamaciones 2
aclamado 2
aclara 2
aclaró 1
activo 1
activos 1
acuchillar -2
acuerdo 1
acusa -2
acusación -2
acusaciones -2
acusado -3
acusado -2
acusados -2
acusando -2
acusar -2
adecuado 1
admira 3
admirado 3
admirar 3
admirativo 3

admite -1
admitir -1
adoctrina -2
adoctrinados -2
adoctrinando -2
adoctrinar -2
adopta 1
adoptar 1
adora 3
adorable 3
adorado 3
adorado 3
adorar 3
adormecer -1

advertencia -3
advertencias -3
advertir -2
advierte -2
afable 2
afectado -1
afecto 3
afición 2
aficionado 2
afilado 1
afirmativamente 2
afligido -2
afligido -1
afligidos -1
afortunado 2
afortunado 3
afrentado -1
agonising -3
agoniza -3
agoniza -3
agonizado -3
agonizado -3

agonizante -3
agonizar -3
agonizar -3
agotado -2
agotado -2
agradable 2
agradable 3
agradable 3
agradecer 2
agradecido 2
agradecido 3
agrava -2
agravado -2
agravante -2
agravar -2
agraviado -2
agresión -2
agresiones -2
agresivo -2
aguante 2
agujero -2

ahogado -2
ahogado -1
ahogarse -2
ahogarse -2
ahorrar 2
airado -3
aislado -1
ajuste 1
alabada 3
alabado 3
alabando 3
alabanza 3
alabanzas 3
alarma -2
alarmado -2

alarmista -2
alarmistas -2
alcance 2
alcanza 1
alcanzado 1
alcanzando 1
alcanzar 1
aleccionador 1
alegre 2
alegres 2
alegremente 3
alegría 3
alegrías 3
alérgico -2
alertar -1
alicaído -2
alienación -2
alienta 2
alivia 1
aliviado 2
aliviando 2
aliviar 1
alterar -2
amable 3
amables 3
amado 3
amada 3
amañado -1

amargado -2
amargamente -2
amargo -2
ambicioso 2
ambivalente -1
amenaza -2
amenazas -2
amenazada -2
amenazado -2
amenazado -2
amenazante -2
amenazar -2
amenazas -2
amenazas -2
amistoso 2

amonestado -2
amonestar -2
amor 3
amores 3
amordazada -2
amoroso 2
amplia 1
ampliar 1
analfabetismo -2
angers -3
angustia -3
angustia -2
angustiado -3
angustiados -2
angustias -2
angustioso -3
anhelo -1
anhelos 1
anima 2
animado 2
animada 2
animar 2
animosidad -2
ansiedad -2
ansioso -2
ansiosa 2
antagónico -2
anticipación 1

antiguo -1
anular -1
apacigua 2
apaciguando 2
apaciguar 2
aparejo -1
apasionado 2
apatía -3
apático -3
apenado -2
apestan -3
aplacado 2
aplastado -2
aplastante -1
aplastar -1

aplaude 2
aplaudido 2
aplaudiendo 2
aplaudir 2
aplausos 2
aplazado -1
aplazar -1
apocado -2
apocalíptico -2
apoya 2
apoyado 2
apoyar 2
apoyo 1
apoyos 2
aprecia 2
aprecio 2
apreciación 2
apreciado 2
apreciada 2
apreciando 2
apreciar 2
aprensivo -2
aprobación 2
aprobado 2
aprobar 2
aprueba 2
aptitud 1
apuñalado -2

ardiente 1
arrastrado -1
arrastrados -1
arrastrar -1
arrepentido -2
arrepentida -1
arrestos -2
arresto -2
arrinconado -2
arrogante -2
arrojado -2
arrojar -1
arruinado -3
arruinados -2
arruinando -2

asco -3
asegura 2
asegurado 2
asegurar 1
aseguró 1
asequible 1
asesinar -3
asesinato -3
asesinado -2
asesinatos -3
asesino -3
asesino -2
asfixia -2
asfixió -2
abochornó -2
asombra 3
asombrado 2
asombrados 2
asombrada 3
asombrar 2
asombroso 3
asombrosa 4
asombroso 5
áspero -2
asuntos 1
asustadizo -2
asustadizos -1
asustado -2

asustada -2
asustar -2
atacado -1
atacando -1
atacar -1
ataques -1
atascado -2
atento 2
atentos 2
aterroriza -3
aterrorizado -3
aterrorizados -3
aterrorizar -3
atracción 2
atracciones 2

atrae 1
atraer 1
atragantó -2
atraído 1
atrapado -2
atrevido 2
atropello -3
aturdido -2
audaz 2
audaces 3
aumenta 1
aumentado 1
aumentar 1
ausente -1
ausentes -1
autor -2
autores -2
autoridad 1
avanzado 1
aventura 2
aventuras 2
aventurero 2
avergonzado -2
avergonzados -2
avergonzada -2
avergonzar -2
avergüenza -2
aversión -2

ávido 2	borrado -2	capacidad 2	chocar -2
ay -1	borrosa -2	capacidades 2	chocho -5
ayuda 2	boyante 2	capaz 1	choque -2
ayudando 2	brillante 1	caprichoso 1	chupar -3
ayudar 2	brillantes 4	carcajada 1	ciego -1
barras -2	brillo 1	carga -2	cielo 2
barrera -2	brilla 3	cargado -2	cierne -1
bastante 1	broma 2	cargas -2	cierto 1
bastardos -5	bruto -2	cargos -2	cínico -2
batalla -1	buena 3	cariñoso 3	cinismo -2
batallas -1	buenas 3	caro -2	claramente 1
batido -2	bueno 3	caros -2	claridad 2
baza 2	bueno 3	castiga -3	claro 1
beatífico 3	buenos 3	castiga -2	clara 1
bebé 2	burla -2	castigado -3	coaccionado -2
bellezas 3	burlado -2	castigada -2	cobarde -2
bendecir 2	burlón -2	castigando -3	codicia -3
bendice 2	burro -4	castigar -3	codicioso -2
bendición 3	cabildeo -2	catástrofe -3	cojo -2
bendiciones 4	cabizbajo -2	catastrófico -4	colapso -2
beneficiado 2	cabreado -4	cauteloso -1	colapsos -2
beneficiando 2	cabrón -5	cautivar 3	colisión -2
beneficio 2	cacharro -1	cautivó 3	colisionando -1
beneficios 2	cadáver -1	celda -2	colisiones -2
beso 2	caer -1	celebra 3	colusión -3
bien 2	caído -2	celebrado 3	combates -1
bienes 2	calambre -1	celebrando 3	combatir -1
bienvenido 2	caliente 1	celebrar 3	comedia 1
bla -2	calma 2	celestial 4	comete 1
blanquear -3	calmaba 3	celoso -2	cometer 1
blockbuster 3	calmado 2	celosa -2	como 2
bloquear -1	calmante 2	censores -2	comodidad 2
bloqueo -1	calmar 3	censores -2	cómodo 2
bloques -1	calor 2	centrado 2	compadecía -1
boba -2	cancela -1	ceñudo -1	compartido 1
boicoteando -2	cancelado -1	cerca 1	compasivo 2
boicotear -2	cancelada -1	engaña -3	competente 2
boicoteó -2	cancelar -1	chispa 1	competitivo 2
boicots -2	cáncer -1	chispas 3	complacer 1
bomba -1	candidez -2	chistes 2	complaciente -2
bondad 3	cansado -2	choca -1	comprometido 1
bonito 2	caos -2	chocante -2	compromiso 2
borracho -2	caótico -2		conceder 1

concedido 1
concesión 1
concilia 2
conciliación 2
conciliado 2
conciliar 2
condena -2
condenación -2
condenado -2
condenar -2
confeccionado -1
confiado 2
confianza 2
confiar 1
conflictiva -2
conflicto -2
conflictos -2
confundido -2
confundiendo -2
confundir -2
confuso -2
congrats 2
coño -5
consentimiento 2
consentimientos 2
consolable 2
consolador 2
conspiración -3
consternada -2
constreñido -2
consuela 2
contagio -2
contagios -2
contagioso -1
contamina -2
contaminado -2
contaminadores -2
contaminar -2
contencioso -2
contender -1
contendiendo -1
contendiente -1
contento 3
contestable -2

contra -1
contradictorio -2
contratar 1
convence 1
convencer 1
convencido 1
convencidos -1
convenido 1
cordialidad 3
corrupción -2
cortar -1
corte -1
cortejar 3
cortes -1
cortés 2
cortesía 2
costoso -2
coziness 2
creados 1
creativo 2
creciente 1
crecimiento 2
crédulo -2
crimen -3
criminal -3
crisis -3
critica -2
crítica -2
criticando -2
criticar -2
criticó -2
crítico -2
críticos -2
cruel -3
crueldad -3
cuchara 3
cueros -1
cuestionable -2
cuestionado -1
cuidadosamente 1
cuidadoso 2
cuidar 2
culo -4
culpa -3

culpabilidad -2
culpable -3
culpado -2
culpando -2
cumple 2
cumplido 2
cumplir 2
curioso 1
dañar -2
daño -3
daña -2
dañas -2
daños -2
débil -2
débiles -2
debilidad -2
decepción -2
decepciona -2
decepcionado -2
decepcionante -2
decepcionar -2
decepciones -2
decisivo 1
dedicado 2
defecto -3
defectos -3
defensor 2
déficit -2
defraudador -4
degrada -2
degradado -2
degradar -2
dejar -1
dejecting -2
dejects -2
deleita 3
deleitando 3
deleitar 3
delincuente -2
delincuentes -3
delito -3
demanda -1
demandando -2
demandas -2

demente -2
demostración -1
denier -2
denuncia -2
denunciar -2
deprimente -2
deprimido -2
derrotado -2
desacreditado -2
desafiante -1
desafiar -1
desagradable -3
desairado -2
desairar -2
desaire -2
desaires -2
desalentado -2
desalojo -1
desanimado -2
desaparece -1
desaparecer -1
desaparecido -2
desaprovechado -2
desasosiego -2
desastre -2
desastres -2
desastroso -3
descalificado -2
descarrila -2
descarriló -2
descartado -1
descartando -1
descartar -1
descartes -1
desconectado -1
desconfiado -3
desconfianza -3
desconsiderado -2
desconsolado -2
desconsuelo -2
descontados -1
descontento -2
descorazonado -2
descreer -2

descuida -2
descuidadas -2
descuidado-2
descuidando -2
desdén -2
desdeñar -2
desdeñosamente -2
desdeñoso -2
deseable 2
deseado 2
desempleo -2
desenfocado -2
deseable 1
deseo 1
deseos 1
deseoso 2
desespera -3
desesperación -3
desesperadamente -3
desesperado -3
desestimación -1
desgarrado -2
deshonesto -2
deshumaniza -2
deshumanizado -2
deshumanizante -2
deshumanizar -2
desigual -1
desilusionado -2
desinformación -2
desmoralizados -2
desmotivado -2
desnonar -2
desorganizado -2
desorientado -2
despectivo -2
despedido -1
despeja 1
despistados -2
despotricar -3
despreciativo -2
desprecio -2
despreocupado -2
destacado 2

destacar 2
desterrar -1
destrozado -2
destrucción -3
destrutivo -3
destruido -3
destruir -3
destruye -3
destruyendo -3
desventaja -2
desventajas -2
desventurado -2
detención -2
detener -2
detener -2
detener -1
detenido -3
deterioradas -2
determinado 2
detesta -3
detestaba -3
detestar -3
deuda -2
devastado -2
devastador -2
devastar -2
diamante 1
dicha 3
diestros 2
difamatorio -2
difícil -1
difiriendo -1
digno 2
dilema -1
dios 1
dipshit -3
discordia -2
disculpa -1
disculpándose -1
disculparse -1
disfraces -1
disfrazado -1
disfrazar -1
disfruta 2

disfrutando 2
disfrutar 2
disfunción -2
disgustado -3
disgusto -2
disparar -1
disparates -2
disparo-2
disputa -2
disputado -2
disputando -2
disputas -2
disruptivo -2
distorsiona -2
distorsionada -2
distorsionando -2
distorsionar-2
distracción -2
distrae -2
distraer -2
distráido -2
disturbios -2
diversión 4
diversiones 3
divertido 3
divertir 3
dolor-2
dolores -2
dolorido -2
doloroso -2
ducha -3
dudaba -1
dudado -1
dudas -1
dudo -1
dudosa -1
dudoso-3
duele -2
dulce 2
duro -1
eficaz 2
eficazmente 2
egoísmo -3
egoísta -3

elación 3
elegante 2
elegante 2
elegantemente 2
elogiar 2
elogió 2
embaucar -2
embellecer 3
embruja -2
emergencia -2
emoción 3
emocionado 5
emocionante 3
empatía 2
empeora -3
empeorado -3
empeoramiento -3
empeorar -3
empequeñecer -2
emprendedor 1
enamoramientos -1
encantada 2
encantado 3
encantador 3
encanto 3
encaprichado 2
encarcelado -2
encendido -2
encubrimiento -3
enemigo -2
enemigos -2
energético 2
enérgico 2
enfermedad -2
enfermedades -2
enfermo -2
enfermos -2
enfriamiento -1
enfurece -2
enfurecer -2
enfurecer -2
enfurecido -2
enfureciendo -2
enfurruñado -2

engaña -3
engañado -3
engañados -2
engañar -3
engaño -3
engañosa -3
engñoso -3
enjuicia -1
enjuiciamiento -1
enjuiciar -1
enloquecedor -3
enmudecido -1
enojado -3
enorme 1
enragés -2

ensordecedor -1
entretenido 2
entristecer -2
entristecidos -2
entusiasta 3
envenenado -2
envidia -1
evidiando -1
envidio -1
envidioso -2
equivocado -2
erróneo -2
error -2
errores -2
erupción -2
escándalo -3
escándalos -3
escandaloso -3
escapar -1
escape -1
escapes -1
escasez -2
escepticismo -2
escéptico -2
escépticos -2
esclavitud -3
esclaviza -2
esclavizados -2

esclavizar -2
escurridizo -2
espalda 1
espantado -2
espantoso -3
espectaculares 4
especulativo -2
espera 2
esperado -1
esperanza 2
esperanzado 2
esperar -1
espinoso -2
espíritu 1
espléndido 3

espumoso 3
esquivando -2
estable 2
estafa -3
estafas -2
estampida -2
estancado -2
estancamiento -2
estereotipo -2
esteriotipado -2
estimación 2
estimado 2
estimula 1
estimulado 1
estimulante 2
estimular 1
estímulo 2
estragos -2
estrangulado -2
estresor -2
estridente -2
estropeado -3
estupendo 4
estúpidamente -2
estúpido -2
ético 2
euforia 3
eufórico 4

evita -1
evitó -1
evitado -1
evitar -1
evitaron -1
exagera -2
exageración -2
exageraciones -2
exagerado -2
exagerando -2
exagerar -2
exaltado 3
exasperado 2
exasperante -2
excelencia 3

excelente 3
excelentes 3
excepcional 5
excitado 3
excitar 3
excluidos -2
excluir -1
exclusión -1
exclusivo 2
excusar -1
exento -1
exhaustivo 2
exigente -1
exigió -1
éxito 2
exitoso 3
exonera 2
exonerado 2
exonerando 2
exonerar 2
expandir 1
exploits -2
exploración 1
exploraciones 1
explotado -2
explotar -2
expone -1
exponer -1

exposición -1
expuesto -1
expulsa -2
expulsado -2
expulsando -2
expulsar -2
extasiar 3
éxtasis 2
extático 4
extrañamente -1
extraño -2
extraño -2
extraño -1
exuberante 4
exultante 3

fabuloso 4
fácil 1
facilidad 2
fallar -2
fallas -2
falseó -2
falsificación -3
falsificaciones -3
falsificado -3
falsificar -3
falta -2
fama 1
famélico -2
fanático -2
fanáticos -2
fantasma -1
fantástico 4
farsa -1
fascina 3
fascinado 3
fascinante 3
fascinar 3
fascista -2
fascistas -2
fatalidad -3
fatiga -2
fatigado -2
fatigoso -2

favorecer 2
favorecido 2
favores 2
favorited 2
favorito 2
Favoritos 2
fe 1
felicidad 3
felicitación 2
felicitaciones 2
felicitar 2
feliz 3
feo -3
fea -3
feos -3
feria 2
ferviente 2
fervoroso 2
festivo 2
fiasco -3
fiel 3
fieles -3
filtrado -1
fingiendo -3
firme 2
fóbico -2
follada -4
fomentar 2
fortalece 2
fortalecer 2
fortalecido 2
fortalecimiento 2
forzado -1
fracasado -2
fracaso -2
fracasos -2
frase -2
frases -2
fraude -4
fraudes -4
fraudulencia -4
fraudulento -4
frenesí -3
frenético -1

fresco 1
friki -2
frotis -2
frustra -2
frustró -2
frustración -2
frustrado -2
frustrando -2
frustrante -2
frustrar -2
fuego -2
fuerte 2
fuerza 2
fuga -1
fumante -2
funeral -1
funerales -1
furioso -3
furtivo -1
fútil 2
galante 3
galantería 3
gallardamente 3
gallito -2
gamberrismo -2
gamberro -2
gana 4
ganado 2
ganador 4
ganancias 2
ganando 2
ganar 2
ganga 2
ganster -3
garantizar 1
garboso 2
gemido -2
gemidos -2
generoso 2
genial 3
gilipollas -4
gimiendo -2
gimió -2
glamour 3

glamouroso 3
gloria 2
glorioso 2
golpe -2
golpeado -1
gracia 1
gracias 2
gracioso 3
gran 3
grande 1
granizo 2
gratificación 2
gratis 1
grave -2
graves -2
gripe -2
gris -1
grises -1
gritar -2
grito -2
gritaron -2
gritó -2
gritos -2
guai 4
guardado 2
guerra -2
guerras -2
gustado 2
gustos 2
hábil 2
habilidades 2
hacha -1
hackeado -1
haha 3
hahaha 3
hambre -2
hambriento -2
harto -3
hermosamente 3
hermoso 3
héroe 2
héroes 2
heroico 3
hilarante 2

hipócrita -2
histeria -3
histérico -3
honesto 2
honor 2
honrado 2
honrada 2
honrando 2
honrar 2
hooligans -2
horrendo -3
horrenda -3
horrible -3
horribles -3
horrorizado -3
horrorizados -2
hosco -2
hostil -2
huelga -2
huelgas -1
huelguista -2
humorous 3
humillación -3
humillado -3
humor 2
hurra 5
huye -1
idiota -3
ignora -1
ignorado -2
ignorancia -2
ignorante -2
ignorar -2
ilegal -3
ilumina 2
iluminando 2
iluminar 2
ilustrado 2
imbécil -3
impaciente -2
impar -2
imparable 2
imparcial 2
impedido -1

imperfecto -2
impermeable 2
impersonal -1
impide -1
implacable -1
impone -1
imponente -1
imponer -1
importancia 2
importante 2
importantes 2
impotente -2
impresiona 3
impresionado 3
impresionante 3
impresionante 4
impresionar 3
impuesto -1
impulsado 1
impulsar 1
inacción -2
inaceptable -2
incapacidad -2
incapacitado -2
incapaz -2
incierto -1
inclinado -2
incomoda -3
incómoda -3
incómodo -2
incompetencia -2
incompetente -2
incomprendido -2
inconexo -2
inconsciente -2
inconveniencia -2
inconveniente -2
incrédulo -1
increíble -1
incumplida -2
indeciso -2
indefenso -2
indeseable -2
indestructible 2

indiferencia -2
indiferente -2
indignación -2
indignado -3
indulgente 1
indultos 2
ineficaz -2
ineficazmente -2
inestable -2
inevitable -1
infantil -2
infatuación 2
infectado -2
infeliz -2
inferior -2
infierno -4
inflamada -2
influyente 2
infortunio -2
infracción -2
infractor -2
ingenuo -2
inhabilitación -1
inhibir -1
injusticia -2
injusto -2
inmortal 2
inmovilizado -1
inmune 1
innova 1
innovación 1
innovador 2
innovar 1
inquieto -2
inquieto -2
inquisición -2
inquisitivo 2
insalubre -2
insatisfecho -2
insatisfecho -2
inseguro -2
insensibilidad -2
insensible -2
insignificante -2

insípido -2
insomnio -2
inspira 2
inspiración 2
inspirado 2
inspirador 2
inspirar 2
insuficiente -2
insultado -2
insultante -2
insultar -2
insultos -2
intacto 2
integridad 2
inteligente 1
inteligente 2
intenso 1
interés 1
interesado 2
interesante 2
intereses 1
interrogado -2
interrogatorio -1
interrumpe -2
interrumpido -2
interrumpiendo -2
interrumpir -2
interrupción -2
interrupciones -2
intimida -2
intimidación -2
intimidado -2
intimidados -2
intitulado 1
intrépido 2
intrigas 1
intrincado 2
inútil -2
inutilidad -2
invencible 2
invitar 1
involucra 1
invulnerable 2
ira -3

iracundo -3
ironía -1
irónico -1
irracional -1
irresistible 2
irresoluto -2
irrespetado -2
irresponsable 2
irreversible -1
irritado -3
irritante -3
irritar -3
jactancioso -2
jauja -2
Jesús 1
jocoso 2
jodeos -3
joder -4
jodete 3
jodido -3
jovial 2
joya 1
joyas 1
júbilo 3
jubiloso 3
juguetón 2
jura -2
jurar -2
justicia 2
justificadamente 2
justificado 2
juvenil 2
ladrón -2
lágrimas -2
lamenta -2
lamentaba -2
lamentable -3
lamentando -2
lamentar -2
lamentó -2
lánguido -2
lástima -2
lastimoso -2
leal 3

lealtad 3
legal 1
legales 1
legalmente 1
lento -2
lesión -2
lesionado -2
letárgico -2
letargo -2
libertad 2
libre 1
limitación -1
limitado -1
límites -1
limpiador 2
limpiar 2
lindo 2
lío -2
listo -2
litigio -1
litigioso -2
lívido -2
llaga -1
llanto -2
llora -2
llorar -1
lluvioso -1
lobby -2
locamente -3
loco -3
locos -2
loca -2
locura -3
locuras -2
logra 2
logrado 2
logrados 2
lol 3
lovelies 3
lucha -2
luchando -2
luchar -2
luchar -1
luchó -2

lúgubre -2
lunático -3
lunáticos -3
luto -2
maduro 2
magnífico 5
mal -3
mala -3
maldecir -1
maldito -4
maldito -4
maldito -3
maldito -2
malditos -4
malditos -4
malentendido -2
malentendiende -2
males -1
malévolo -2
malhumorado -2
malinterpretado -2
malisima -4
malo -3
malos -3
malvado -2
mamón -4
mamónes -4
mancha 2
manifestantes -2
manipulación -1
manipulado -1
maravilla 3
maravillas 3
maravilloso 3
maravillosa 4
maravillosos 4
marca 2
mareado -2
mareada -1
maricón -3
maricones -3
masificada -3
masificado -3
mata -3

matanza -3
matar -3
materia 1
mayor 3
mayores 3
meando -3
mear -4
medalla 3
mediocridad -3
meditativo 1
mejor 3
mejora 2
mejorado 2
mejorar 2
mejores 3
melancolía -2
melancólico -1
menospreciado -2
menospreciar -2
menosprecie-2
mentira -3
mentiroso -3
mentirosos -3
mercachifle -2
metódico 2
miedo -2
miedoso 2
mierda -4
mierdoso -4
milagro 4
minado -2
minar -2
miope -2
miopía -2
miserable -3
miseria -2
misericordia 2
misterioso -2
mito -1
moda -2
modernos 1
molesta -2
molesto -2
molestado -2

molestar -2
molestia -2
monopoliza -2
monopolizada -2
monopolizando -2
monopolizar -2
moping -1
mordaz -3
mordaza -2
morir -3
mortificado -2
motín -2
motivación 1
motivado 2
motivar 2
mudo -3
muerte -2
muerto -3
multa 2
murió -3
musaraña -4
natural 1
nafragio -2
necesidades -2
necesitado -2
negado -2
negadores -2
negando -2
negándose -2
negar -2
negatividad -2
negativo -2
negligencia -2
negro -5
nervios -1
nerviosamente -2
nervioso -2
nerviosos -2
nerviosa -2
niega -2
no -2
noble 2
nostálgico -2
notable 2

notorio -2
 novela 2
 novelas 2
 nubló -1
 nueces -3
 obligado 1
 obligatorio -1
 obliterar -2
 obsceno -2
 obsesionado 2
 obsesionante 1
 obsesionar -1
 obsoleto -2
 obstáculo -2
 obstáculos -2

 obstinado -2
 obstruido -1
 ocultación -1
 ocultado -1
 ocultar -1
 odia -3
 odiado -3
 odiando -3
 odio -3
 odia -3
 odioso -3
 ofende -2
 ofender -2
 ofendido -2
 oks 2
 olvidadizo -2
 olvidado -1
 olvidar -1
 oportunidad 2
 oportunidades 2
 opresivo -2
 oprimidos -2
 optimismo 2
 optimista 2
 ora 1
 orando 1
 orar 1
 orgulloso 2

oscuridad -1
 oscuridad -1
 otorgado 3
 oxímoron -1
 pacíficamente 2
 pacífico 2
 pagar -1
 paliza -1
 pánico -3
 pánicos -3
 parada -1
 paradoja -1
 paraíso 3
 parcial -2
 parcialidad -1

 parlamentar -1
 parodia -2
 paros -2
 participación 1
 partidario 1
 partidarios 1
 pasivamente -1
 pasivo -1
 pasmosamente 3
 patético -2
 paz 2
 pecaminoso -3
 peligro -2
 pena -2
 penosa -2
 penoso -2
 pensativo -1
 pensativo 2
 peor -3
 peores -3
 perdedor -3
 perder -3
 perdedor -2
 perdición -2
 pérdida -3
 perdido -3
 perdido -2
 perdiendo -2

perdón 2
 perdonado 2
 perdonador 1
 perdonando 2
 perdonar 1
 perezoso -1
 perfecciona 2
 perfeccionado 2
 perfectamente 3
 perfecto 3
 perforado -2
 perjudicado -2
 perjudicados -2
 perjudicial -2
 perjurio -3

 permitir 1
 perplejo -2
 perpleja -2
 perra -5
 perras -5
 persecución -2
 perseguidos -2
 perseguir -2
 persigue -2
 perspectiva 1
 perspectivas 1
 perturba -2
 perturbado -2
 pésima -3
 pesimismo -2
 pesimista -2
 petrificada -2
 picado -2
 picante -2
 pierde -3
 pinchazo -5
 pintoresco 2
 pique -2
 pistola -1
 placer 3
 pleito -2
 pobre -2
 pobreza -1

pobrisima -2
 polémica -2
 polémico -2
 polla -5
 popular 3
 posesivo -2
 posibilidades 2
 positivo 2
 pospone -1
 posponer -1
 potente 2
 precioso 3
 premio 3
 premios 3
 preocupado -3

 preocupada -2
 preocupante -3
 preocupantes -2
 preocuparse -3
 preparado 1
 presar -2
 presión -1
 presionado -2
 preso -2
 presos -2
 prestigio 3
 pretende -1
 pretender -1
 pretendiendo -1
 prevenido -2
 prisión -2
 privación -2
 privilegiado 2
 proactivo 2
 problema -2
 problemas -2
 problematico -2
 procesado -2
 progreso 2
 prohibido -2
 prohibir -2
 promesa 1
 promete 1

prometido 1
promociona -2
promover 1
promovido 1
promueve 1
propaganda -2
próspero 3
protege 1
proteger 1
protegido 1
protesta -2
protestando -2
protestas -2
provechoso 2
provoca -1
provocado -1
provocando -1
provocar -1
pseudociencia -3
puesto -2
puñalada -2
puñaladas -2
punitivo -2
pusilánimes -2
puta -5
putas -4
queda -2
quejarse -2
querer 1
querido 2
rabia -2
rabioso -2
racismo -3
racista -3
racistas -3
ranter -3
ranters -3
rants -3
rápidamente 2
rápido 2
raptados 2
rapto 2
ratificado 2
realizable 1

realizables 1
rebelión -2
recelo -2
recesión -2
rechaza -1
rechazado -1
rechazados -2
rechazar -2
recomendado 2
recomendar 2
recomienda 2
recompensado 2
recompensar 2
recompensas 2
recto 1
redimido 2
redujo -2
refugios -1
regalo 2
regaño -2
regocija 3
regocijado 3
regocijarse 4
reír 1
reivindica 2
relajado 2
relajante 2
relax 2
remordimiento -2
renuncia -1
renunciar -1
repugnante -3
rescatado 2
rescatar 2
rescate -2
rescates 2
resentido -2
residuos -1
resignado -1
resistente 2
resolver 2
respalda 2
respaldado 1
respetado 2

responsable 2
restaura 1
restauración 1
restaurado 1
restaurar 1
restregones -3
restricción -2
restringe -2
restringido -2
restringiendo -2
restringir -2
resucitar 2
resuelto 1
resuelto 2
resuelve 1
resuelven 2
retardado -2
retardar -2
retenido -1
retirada -3
retirarse -1
retozón 2
retrasado -1
retrasar -1
retraso -1
retrasos -2
revendedor -2
revendedores -2
revive 2
rezagado -2
rico 2
ridiculizado -2
ridiculizando -2
ridiculizar -2
ridículo -3
riendo 1
riesgo -2
riesgos -2
rigurosamente 3
riguroso 3
rió 1
riqueza 3
roba -2
robado -2

robar -2
robó -2
robaron -4
robing -2
robo -4
robos -2
robusto 2
rofl 4
rollo -2
rompió -1
roscado -2
roto -1
roza -2
ruidosa -3
ruidoso -1
ruina -2
ruinas -2
rumiando -2
ruptura -2
saboreando 2
sabotear -2
sabroso 3
saliente 1
saluda 1
saludable 2
saludar 1
saludo 1
saludó 1
saludos 2
salvavidas 4
sangriento -3
sarcástico -2
satisfecho 2
satisfechos 3
scumbag -4
secuestrados -2
secuestro -2
secuestros -2
sedición -2
sedicioso -2
seducido -1
seguridad 1
seguro 2
sensación 1

sensible 2
sensiblero -1
sentencia -2
sentido 3
separado -1
sereno 2
serio 2
sexy 3
shocks -2
sí 1
si 1
significado 1
significativo 2
sii 2
silenciar -1
simpatía 2
simpático 2
simplificación -2
simplificada -2
sinceramente 2
sinceridad 2
sincero 2
siniestro 3
sobornar -3
sobrecarga -1
sobrecargando -2
sobrepasado -2
sobrepeso -1
sobrereaccionar -2
sobrerreacción -2
sobrerreaccionado -2
sobresaltado -2
sobrevendido -2
sobreviviente 2
sobrevivió 2
socava -2
socavando -2
sofisticado 2
sol 2
solemne -1
solidaridad 2
sólido 2
solitario -2
solo -2

solución 1
soluciones 1
sombrió -2
son-of-a-bitch -5
sonreír 2
sonriente 2
sonrió 2
sonrisas 2
sorprende 2
sorprendido -2
sosegado 2
sospecha -1
sospechar -1
sospechar -1
sospechosos -1
spammer -3
spammers -3
spamming -2
subestimaciones -1
subestimado -1
subestimar -1
subrayó -2
subvenciones 1
subversivo -2
sucia -3
suciedad -2
sucio -2
suelto -3
sueño 1
sueños 1
suerte 3
sufre -2
sufrimiento -2
sufrir -2
suicida -2
suicidio -2
súper 3
superchería -2
superior 2
suspender -1
suspendido -1
suspicaz -2
suspiro -2
sustancial 1

sustancialmente 1
susto -2
tacaño -2
tambaleante -2
tard -2
tedio -2
telar -1
temblante -2
temblor -2
temerario -2
temeroso -2
temía -2
temible -2
temiendo -2
tensión -1
tenso -2
tercermundista -3
tergiversación -2
terrible -3
terriblemente -3
terror -3
tesoro 2
tesoros 2
teta -2
tierno 2
tímido -2
tímida -1
timorato -2
tipo 2
tirón -3
tolerante 2
tonto -1
top 2
tops 2
torpe -2
tortura -4
torturado -4
torturar -4
torturas -4
totalitario -2
totalitarismo -2
tóxico -2
tragedia -2
trágico -2

traición -3
traiciona -3
traicionado -3
traicionando -3
traicionar -3
traidor -3
tramado -2
trampa -1
tramposo -3
tramposos -3
tranquila 2
tranquiliza 1
tranquilizador 2
tranquilizar 1
tranquilo 2
trastorna -2
trastorno -2
trauma -3
traumático -3
travesura -1
trémulo -2
triste -2
tristemente -2
triumfante 4
triunfo 4
tumor -2
unido 1
unificado 1
unirse 1
urgente -1
usado -1
útil 2
útiles 2
utilidad 2
vacilando -1
vacilante -2
vacilar -2
vacío -1
vacíos -1
vago -2
válida 1
validado 1
validar 1
valiente 2

valientemente 2
valor 2
vecinos -1
vejación -2
venenos -2
venerado 2
venganza -2
vengativo -2
ventaja 2
ventajas 2
ventilador 3
veredicto -1
veredictos -1
vergonzoso -2
vergonzosa -2
vergüenza -2

vergüenzas -2
vertederos -1
vestido -2
vibrante 3
vicioso -2
víctima -2
víctimas -3
victimiza-3
victimizar -3
vieja -1
vigilante 3
vil -3
vindicada 2
vindicando 2
vindicar 2
viola -2

violación -4
violado -2
violador -4
violar -2
violencia -3
violento -3
virtuoso 2
virulento -2
visión 1
visionario 1
visionarios 3
visiones 1
vitalidad 3
vitamina 1
vitoreó 2
vitriólico -3

viudo -1
vivaz 3
vivo 1
vociferante -1
voluntad 2
vulnerabilidad -2
vulnerable -2
winwin 3
won 3
FALSO -2
VERDADERO 2
falsos -2
verdaderos 2

Diccionario del sector bancario

BBVA

Banc

banco

Bco

Kxa

acciones

acciones de bolsa

ahorro

ampliacion de hipoteca

ampliacion hipoteca

ampliar hipoteca

aprender a invertir

banc

banca

banca de inversion

banca etica

banca on line

banca online

banca santander

bancaonline

banco

banco andalucia

banco azteca

banco bbva

banco bbva chile

banco bci

banco bicentenario

banco bilbao vizcaya

banco castilla

banco central de chile

banco chile

banco citibank

banco ciudad

banco corpbanca

banco credichile

banco credicoop

banco de bogota

banco de chile

banco de galicia

banco de inversion

banco de la nacion

banco de santander

banco de venezuela

banco del estado

banco del estado de chile

banco edwards

banco espana

banco espirito santo

banco estado

banco falabella

banco frances

banco galicia

banco hipotecario

banco hsbc

banco industrial

banco internacional

banco itau chile

banco macro

banco mercantil

banco nacion

banco online

banco pastor

banco patagonia

banco popular

banco popular depositos

banco popular e

banco popular espanol

banco popular particulares

banco primus

banco provincia

banco provincia de buenos aires

banco provincial

banco rio

banco santader

banco santander

banco santander central hispano

banco santander chile

banco santander depositos

banco santander hipotecas

banco santander nomina

banco santander rio

banco santander santiago

banco satander

banco scotiabank

banco security

bancochile

bancomer	bolsa valores	calculo de hipoteca
bancos	bolsas de empleo	calculo de hipotecas
bancos de inversion	bolsas de trabajo	calculo de prestamos
bancos online	bovespa	calculo hipoteca
bancos y cajas	buro de credito	calculo hipotecas
bancosantander	buscador de fondos de inversion	calculo intereses deposito
banesto depositos	buscador de hipotecas	calculo pension jubilacion
banesto hipotecas	buscador hipotecas	calculo prestamo
banesto nomina	busco empleo	cambiar de hipoteca
bankia depositos	caja madrid	cambiar hipoteca
bankia hipotecas	caja rural	cambiar hipoteca de banco
bankimia depositos	cajamadrid	cambio de hipoteca
bankinter cuenta nomina	calcula hipoteca	cambio hipoteca
bankinter depositos	calcula tu hipoteca	carta de credito
bankinter planes de pensiones	calculador de hipotecas	cofidis prestamos
barclays bank	calculador hipotecas	como ahorrar dinero
bbva	calculadora de hipotecas	como conseguir dinero
bbva cuenta nomina	calculadora de prestamos	como hacer nominas
bbva depositos	calculadora hipoteca	como hacer una nomina
bbva depositos a plazo fijo	calculadora hipotecaria	como invertir
bbva domiciliar nomina	calcular cuota hipoteca	como invertir dinero
bbva fondos de inversion	calcular hipoteca	como invertir en bolsa
bbva nomina	calcular hipoteca bbva	como invertir en la bolsa
bbva planes de pensiones	calcular hipoteca simulador	como invertir en oro
bco	calcular hipotecas	como invertir mi dinero
bco santander	calcular intereses plazo fijo	como operar en bolsa
bolsa	calcular nomina	como se hace una nomina
bolsa de	calcular prestamo hipotecario	comparador de depositos
bolsa de trabajo	calcular prestamo personal	comparador de depositos bancarios
bolsa de valores	calcular prestamos personales	comparador de fondos de inversion
bolsa online	calcular una hipoteca	comparador de hipotecas
bolsa trabajo	calculo cuota hipoteca	comparador de planes de pensiones

comparador depositos
comparador depositos bancarios
comparador hipotecas
comparador planes de pensiones
comparar hipotecas
comparativa cuentas nomina
comparativa de depositos
comparativa de hipotecas
comparativa depositos
comparativa depositos bancarios
comparativa fondos de inversion
comparativa hipotecas
comparativa planes de pensiones
comprar acciones
comprar con tarjeta de credito
cotizacion fondos de inversion
cotizaciones bolsa
credifacil
credito
credito bancario
credito con asnef
credito facil
credito hipotecario
credito infonavit
credito inmediato
credito online
credito personal
credito rapido
credito si
credito ya
creditos
creditos bancarios

creditos con asnef
creditos de consumo
creditos economicos
creditos hipotecarios
creditos inmediatos
creditos on line
creditos online
creditos online inmediatos
creditos online rapidos
creditos personales
creditos personales rapidos
creditos rapidos
creditos rapidos online
creditos sin aval
creditos sin nomina
creditos urgentes
cuenta corriente
cuenta naranja
cuenta nomina
cuenta nomina banesto
cuenta nomina bankinter
cuenta nomina bbva
cuenta nomina ing
cuenta nomina regalos
cuenta nomina santander
cuentas nomina
cuentas nomina regalos
cuentas nominas
cuota hipoteca
curso bolsa
deposito
deposito 4

deposito a plazo fijo
deposito azul
deposito azul ibanesto
deposito bancario
deposito banco popular
deposito banco santander
deposito bankia
deposito bbva
deposito caixa
deposito gasol
deposito gasol banco popular
deposito ing
deposito la caixa
deposito plazo fijo
deposito santander
depositos a 6 meses
depositos a plazo
depositos a plazo fijo
depositos a plazo fijo bbva
depositos a un año
depositos alta rentabilidad
depositos bancarios
depositos bancarios mas rentables
depositos banco popular
depositos banco santander
depositos bancos
depositos banesto
depositos bankia
depositos bankinter
depositos bbva
depositos caixa
depositos caja madrid

depositos cajamar	donde invertir 10000 euros	fondos de inversion ranking
depositos de la caixa	donde invertir dinero	fondos de inversion renta fija
depositos estructurados	donde invertir en bolsa	fondos de inversion santander
depositos ibercaja	donde invertir mi dinero	fondos de inversiones
depositos la caixa	ejemplo nomina	fondos de inversión
depositos mas rentables	el mejor deposito	fondos de pensiones
depositos online	el mejor fondo de inversion	fondos garantizados
depositos plazo	el mejor plan de pensiones	fondos inversion
depositos plazo fijo	en que invertir	fondos inversion garantizados
depositos rentables	en que invertir dinero	fondos inversion santander
depositos santander	en que invertir mi dinero	franquicias rentables
depósitos	entidades financieras	gastos hipoteca
desgravacion plan de pensiones	estrategias de inversion	gestoras de fondos de inversion
deudas	financiacion	hipoteca
dinero a plazo fijo	financiamiento	hipoteca 100
dinero plazo fijo	financiera	hipoteca 100 mas gastos
dinero rapido	financieras	hipoteca azul
dinero urgente	financieras prestamos	hipoteca banco popular
domicilia tu nomina	financieras prestamos personales	hipoteca banesto
domiciliacion de nomina	fiscalidad fondos de inversion	hipoteca barata
domiciliacion nomina	fiscalidad planes de pensiones	hipoteca bbva
domiciliacion nomina bbva	fondo de inversion	hipoteca caixa
domiciliar la nomina	fondo de inversion garantizado	hipoteca cambio de casa
domiciliar nomina	fondo de inversiones	hipoteca ing
domiciliar nomina banesto	fondo de pensiones	hipoteca ing direct
domiciliar nomina bbva	fondo inversion	hipoteca inversa
domiciliar nomina en bbva	fondos de inversion banco santander	hipoteca joven
domiciliar nomina regalo	fondos de inversion bbva	hipoteca la caixa
domiciliar nomina regalos	fondos de inversion de ibercaja	hipoteca multidivisa
domiciliar nomina santander	fondos de inversion garantizados	hipoteca naranja
donde domiciliar la nomina	fondos de inversion la caixa	hipoteca puente
donde invertir	fondos de inversion mas rentables	hipoteca santander

hipoteca segunda vivienda	intereses a plazo fijo	invertir en dolares
hipoteca simulador	intereses bancarios	invertir en fondos
hipotecas	intereses depositos	invertir en fondos de inversion
hipotecas 100	intereses depositos bancarios	invertir en forex
hipotecas 100 financiacion	intereses plazo fijo	invertir en oro
hipotecas 100 mas gastos	inverco planes de pensiones	invitacion de boda
hipotecas al 100	inversion	invitaciones
hipotecas banco popular	inversion en bolsa	invitaciones boda
hipotecas banesto	inversion rentable	invitaciones de bautizo
hipotecas baratas	inversion segura	invitaciones de boda
hipotecas bbk	inversiones	invitaciones de boda baratas
hipotecas bbva	inversiones a plazo fijo	invitaciones de boda gratis
hipotecas la caixa	inversiones en bolsa	invitaciones de boda originales
hipotecas mas baratas	inversiones financieras	invitaciones de bodas
hipotecas online	inversiones financieras a corto plazo	invitaciones de comunion
hipotecas santander	inversiones financieras temporales	invitaciones para imprimir
hipotecas y euribor	inversiones inmobiliarias	jubilacion
ibanesto hipoteca	inversiones rentables	la bolsa de valores
ibercaja	inversiones rentables a corto plazo	La Caixa
ibercaja depositos	inversiones seguras	la caixa depósitos
ibercaja fondos de inversion	inversiones seguras y rentables	la mejor cuenta nomina
ibex 35	inversionistas	la mejor hipoteca
imposicion a plazo fijo	inversis banco	la mejor inversion
imposicion plazo fijo	invertir	la mejor tarjeta de credito
imposiciones a plazo fijo	invertir bolsa	la tarjeta de credito
infobolsa	invertir dinero	las mejores cuentas nomina
ing cuenta nomina	invertir dinero seguro	las mejores hipotecas
ing direct cuenta nomina	invertir en	las mejores inversiones
interes a plazo fijo	invertir en acciones	las mejores tarjetas de credito
interes depositos	invertir en bolsa	las tarjetas de credito
interes hipoteca	invertir en bonos	ley planes de pensiones
interes plazo fijo	invertir en divisas	linea de credito

los fondos de inversion	mejores intereses bancarios	numero tarjeta de credito
los mejores depositos	mejores inversiones	numeros de tarjetas de credito
los mejores depositos bancarios	mejores planes de pensiones	nómina
los mejores fondos de inversion	mejores planes de pensiones garantizados	oferta domiciliar nomina
los mejores planes de pensiones	mejores plazos fijos	oferta nomina
los mejores plazos fijos	mejores prestamos personales	ofertas bancarias
mapfre fondos de inversion	mejores tarjetas de credito	ofertas bancos nomina
mapfre planes de pensiones	mejores depositos	ofertas depositos bancarios
mastercard	mercado de valores	ofertas domiciliacion nomina
mediolanum	microcredito	ofertas domiciliar nomina
mejor banco para domiciliar nomina	microcreditos	ofertas nomina
mejor cuenta nomina	microcreditos rapidos	ofertas planes de pensiones
mejor deposito	mini creditos rapidos	ofertas por domiciliar la nomina
mejor deposito bancario	mini prestamos rapidos	ofertas por domiciliar nomina
mejor fondo de inversion	minicreditos sin nomina	operar en bolsa
mejor fondo inversion	mutuo	pago con tarjeta de credito
mejor hipoteca	necesito dinero	pedir credito
mejor inversion	necesito dinero urgente	pedir prestamo
mejor plan de pensiones	necesito prestamo	pedir prestamo personal
mejor plazo fijo	necesito un prestamo	pedir tarjeta de credito
mejor tarjeta de credito	negocios con poca inversion	pension
mejores cuentas nomina	negocios rentables	pension de jubilacion
mejores cuentas nominas	negocios sin inversion	pensiones
mejores depositos	nomina banco	pisos de bancos
mejores depositos a plazo	nomina bbva	plan de ahorro
mejores depositos a plazo fijo	nomina electronica	plan de jubilacion
mejores depositos bancarios	nomina ipad	plan de pension
mejores depositos plazo fijo	nomina regalo	plan de pensiones
mejores fondos de inversion	nomina santander	plan de pensiones bbva
mejores fondos de inversion garantizados	nominas	plan de pensiones caja madrid
mejores hipotecas	novacion hipoteca	plan de pensiones de telefonica
mejores hipotecas del mercado	numero de tarjeta visa	plan de pensiones garantizado

plan de pensiones ing	prestamistas de dinero	prestamos online inmediatos
plan de pensiones la caixa	prestamistas particulares	prestamos para empresas
plan de pensiones mapfre	prestamistas privados	prestamos personal
plan de pensiones santander	prestamo	prestamos personales
plan de prevision asegurado	prestamo bancario	prestamos personales bbva
plan jubilacion	prestamo coche	prestamos personales con asnef
plan pensiones	prestamo de dinero	prestamos personales on line
plan pensiones bbva	prestamo dinero	prestamos personales online
plan pensiones santander	prestamo hipotecario	prestamos personales rapidos
plan prevision asegurado	prestamo inmediato	prestamos personales santander
planes de jubilacion	prestamo online	prestamos personales sin aval
planes de pension	prestamo personal	prestamos personales urgentes
planes de pensiones	prestamo personal bbva	prestamos por internet
planes de pensiones bbva	prestamo personales	prestamos privados
planes de pensiones garantizados	prestamo rapido	prestamos rapidos
planes de pensiones la caixa	prestamo urgente	prestamos rapidos con asnef
planes de pensiones mapfre	prestamos al instante	prestamos rapidos online
planes de pensiones mas rentables	prestamos bancarios	prestamos rapidos sin aval
planes de pensiones santander	prestamos bbva	prestamos rapidos sin nomina
planes pensiones	prestamos con asnef	prestamos santander
plans de pensions	prestamos con garantia hipotecaria	prestamos sin aval
plazo fijo	prestamos de dinero	prestamos sin nomina
plazo fijo bbva	prestamos de dinero rapido	prestamos urgentes
plazo fijo santander	prestamos dinero	productos de inversion
plazos fijos	prestamos en efectivo	promocion nomina
plazos fijos mas rentables	prestamos faciles	promociones nomina
popular banco	prestamos hipotecarios	préstamos
postales	prestamos hipotecarios bbva	quiero invertir
postales gratis	prestamos ico	rankia depositos
postales virtuales	prestamos inmediatos	ranking de fondos de inversion
prestamista	prestamos on line	ranking depositos
prestamistas	prestamos online	ranking fondos de inversion

ranking planes de pensiones	santander online	simulador prestamos personales
recuperar plan de pensiones	satander	simulador rescate plan de pensiones
refinanciacion	segunda hipoteca	simuladores de hipotecas
reforma de las pensiones	simulacion de hipoteca	simuladores hipotecas
regalo cuenta nomina	simulacion hipoteca	simular hipoteca
regalo domiciliario nomina	simulacion hipotecas	sobres
regalo nomina	simulador credito	solicitar credito
regalo por domiciliario nomina	simulador credito hipotecario	solicitar prestamo personal
regalos cuenta nomina	simulador de bolsa	solicitar tarjeta credito
regalos domiciliario nomina	simulador de credito	solicitar tarjeta de credito
regalos nomina	simulador de creditos	solicitar tarjeta de credito online
regalos nominas	simulador de depositos	solicitar tarjeta visa
regalos por domiciliario la nomina	simulador de hipoteca	solicitar tarjetas de credito
regalos por domiciliario nomina	simulador de hipotecas	solicitar tarjetas de credito online
regalos por nomina	simulador de hipotecas bbva	solicitar una tarjeta de credito
rentabilidad	simulador de prestamos	solicitud de tarjeta de credito
rentabilidad depositos	simulador de prestamos hipotecarios	solicitud tarjeta credito
rentabilidad fondos de inversion	simulador de prestamos personales	solicitud tarjeta de credito
rentabilidad planes de pensiones	simulador depositos	subrogacion de hipoteca
rescatar plan de pensiones	simulador depositos bancarios	subrogacion hipoteca
rescate de plan de pensiones	simulador fondos de inversion	subrogacion hipoteca sin gastos
rescate plan de pensiones	simulador hipoteca	subrogacion hipotecas
rescate plan pensiones	simulador hipoteca bbva	subrogar hipoteca
reunificacion de deudas	simulador hipotecario	suelo hipotecas
reunificacion de prestamos	simulador hipotecas	targeta credito
revista inversion	simulador plan de pensiones	targeta de credito
santader	simulador planes de pensiones	targetas credito
santande	simulador plazo fijo	targetas de credito
santander	simulador prestamo	tarjeta american express
santander banco	simulador prestamo hipotecario	tarjeta bancaria
santander depositos	simulador prestamo personal	tarjeta citibank
santander nomina	simulador prestamos hipotecarios	tarjeta credito

tarjeta credito gratis
tarjeta credito gratuita
tarjeta credito online
tarjeta credito virtual
tarjeta de credito bbva
tarjeta de credito citibank
tarjeta de credito debito
tarjeta de credito gratis
tarjeta de credito gratuita
tarjeta de credito la caixa
tarjeta de credito mastercard
tarjeta de credito o debito
tarjeta de credito online
tarjeta de credito paypal
tarjeta de credito prepagada
tarjeta de credito sin nomina
tarjeta de credito virtual
tarjeta de credito visa
tarjeta de credito y debito
tarjeta de crédito

tarjeta de debito
tarjeta debito
tarjeta mas
tarjeta mastercard
tarjeta mastercard prepago
tarjeta naranja
tarjeta visa
tarjetas
tarjetas bancarias
tarjetas comerciales
tarjetas credito
tarjetas credito gratis
tarjetas credito online
tarjetas creditos
tarjetas de credito
tarjetas de credito bbva
tarjetas de credito gratis
tarjetas de credito gratuitas
tarjetas de credito la caixa
tarjetas de credito on line

tarjetas de credito para empresas
tarjetas de credito rapidas
tarjetas de credito sin nomina
tarjetas de credito visa
tarjetas de credito y debito
tarjetas de debito
tarjetas visa
tipo de interes
tipos de fondos de inversion
tipos de hipotecas
tipos de interes
tipos de interes hipotecas
tipos de tarjetas de credito
trae tu nomina
traspasar plan de pensiones
traspaso plan de pensiones
tu capital
unicaja nomina
visa
visa electron

Transcripción de las entrevistas

Transcripción de las entrevistas en profundidad realizadas a directores/as de comunicación digital de empresas tanto del sector de gran consumo del sector financiero para desarrollar el punto cinco de la tesis, referente a la investigación cualitativa.

Entrevista: e-LaCaixa

Nombre: Andreu Sans

Cargo: Responsable Digital Networks

1. *¿Cuál es el papel de las redes sociales en su empresa?*

Representan una oportunidad de disponer de un canal más de relación con los clientes. Forma parte de la estrategia multicanal de la empresa, dado que los clientes ya son multicanal.

Es una oportunidad para conectar servicios a través de las redes sociales.

Es una presencia corporativa junto con una presencia especializada. Por ej. Acceso a L.O. desde Facebook. O un servicio de geolocalización de oficinas a través de Twitter.

Entre el 20% al 30% de los clientes proceden de las redes sociales. Una relación online que se desarrolla gracias a propuestas de alto valor añadido.

2. *¿Por qué su empresa está presente en las redes sociales?*

Porque los clientes ya están presentes y utilizan habitualmente las redes sociales.

eLaCaixa además de estar presente en las más importantes redes sociales. Ha desarrollado redes sociales propias y exclusivas para segmentos específicos de clientes como:

- a) Caixa Negocis Community: (>Oct2010) con unos 60.000 usuarios
- b) Xarxa Club Ara para > 65 años

- c) Inspiran's (>Jun2012): una ventana abierta a la aportación directa de ideas para mejorar servicios
- d) Stock tactics (>nov2013) con unos 10.000 usuarios para intercambiar experiencias y conocimiento alrededor del mercado de valores.

3. *¿Cuáles son los principales beneficios que espera conseguir con su presencia en las redes sociales?*

- a) Ofrecer servicios de valor añadido a los clientes.
- b) Obtener información, inputs y más conocimiento del cliente
- c) Facilita un acceso mejor o más cualitativo a los clientes. Nuevas formas de comunicación, como las infografías.
- d) La captación y fidelización de clientes queda implícita a través de las propuestas de servicios de alto valor.

4. *¿En qué redes sociales su empresa está presente (tiene una o varias cuentas creadas)?*

- Facebook, Twitter, Youtube, LinkedIn, Google +, Instagram,
- Pinterest, Tuenti, __Todas las anteriores + el Blog

5. *¿Cuál de las redes sociales considera más importante para su empresa? (mencionar sólo una)*

- Facebook, Twitter, Youtube, LinkedIn, Google +, Instagram,
- Pinterest, Tuenti, _____

6. *¿Cómo se realiza la gestión de las redes sociales? (internamente, subcontratada, parcialmente subcontratada)*

Interna, con una gestión transversal entre todos los departamentos.

Proveedores colaboran en la generación de contenidos, diseño y parte de la monitorización.

7. *Dentro de la empresa, ¿Qué personas son las que gestionan las redes sociales?*

Gestión transversal en función del contenido.

Ej: SAC: la relación atención al cliente lo gestiona el departamento de Contact Center.

La gestión de las redes sociales privadas de la Caixa, la realiza el departamento interno de Digital Networks.

No existe la figura del responsable de una red social.

Si existe un comité coordinador de las redes sociales para todos los departamentos.

8. *¿De qué departamento depende este equipo de personas?*

De cada departamento

9. *¿Monitorizan la presencia en las redes sociales?*

Si. Casi en su totalidad de una forma interna.

10. *¿Qué herramientas de monitorización utilizan para el seguimiento de su presencia en las redes sociales?*

La herramienta mencionada es : BrandWatch

Han desarrollado herramientas internas de monitorización

Han ido cambiando a medida que han probado nuevas herramientas que han ido apareciendo en un mercado muy cambiante.

11. *¿Con qué frecuencia emiten informes de seguimiento de la actividad en las redes sociales?*

Existen varias categorías de informes, las cuales determinan su frecuencia.

Hay alertas diarias e informes semanales, mensuales y anuales en función del tema e interés.

12. *¿Qué métricas consideran como las más importantes a monitorizar?*

- Variables cuantitativas como: audiencias, coberturas, actividad en las redes sociales
- Variables cualitativas: estudios internos con interpretación por parte de personal experimentado.

13. *¿Han utilizado alguna vez, las herramientas de monitorización para identificar posibles insights?*

La red social privada Inspiran's, ya es una fuente de obtención sistematizada de información de los clientes. Actitud activa por parte de la Caixa para solicitar información directamente a los clientes.

Se analizan tendencias y se obtienen oportunidades.

A partir de clientes que recomendaban productos se decidió entregar la Ciber Tarjeta en el pack básico de alta de cliente online.

14. *¿Monitorizan la métrica "Sentiment" (valora los comentarios positivos, negativos o neutros) que realizan sus consumidores?*

La variable "Sentiment" la siguen. Da mejores resultados en inglés que en castellano o catalán.

Sirve como termómetro en líneas generales, pero por su falta de precisión (no interpreta bien las ironías, ...) no se analiza puntualmente.

Pensando en el futuro considera que se tiene que desarrollar mucho más para ser más precisa y fiable.

Considera que las herramientas de monitorización de los sentimientos/emociones han de evolucionar.

Considera que estas herramientas, lo más importante que pueden ofrecer es la capacidad para resumir lo más relevante de entre todo el ruido: “ De qué se habla”.

Entrevista Banco Sabadell-Atlántico

Nombre: Ferran Pallàs Piñana

Departamento: Dirección de Marketing Digital

Cargo: Especialista en Social Media Marketing

1. *¿Cuál es el papel de las redes sociales en su empresa?*

Inicio 2010. Desde SAC también a través de las RS.

Ahora ha crecido mucho. Dividido en 3 partes

1.- SAC

2.- Marketing y Publicidad:

adaptación de campañas publicitarias del off-line al on-line o bien crear campañas propias específicas para el entorno de las RS.

3.- Comunicación relación con los medios: periodistas, prensa,.....

2. *¿Por qué su empresa está presente en las redes sociales?*

Comunicación, Seguridad y Confianza

Para establecer relaciones con los clientes a largo plazo.

Seguridad: que el usuario sepa que el banco siempre responde antes sus dudas

Confianza: que hay alguien detras, un equipo de personas a tu servicio.

3. *¿Cuáles son los principales beneficios que espera conseguir con su presencia en las redes sociales?*

1. Sensación real de proximidad. Qué piensa el usuario de primera mano.

2. Alcanzar también a los potenciales clientes de forma más transparente, transversal y real.

La mejor forma de captación es a través de la fidelización.

Es un gran canal para obtener datos y sacar a los clientes de su entorno: facebook o Tw y ponerlos en un entorno propio de la marca para rellenar un formulario.

Tracking total desde el origen para saber el roi de las inversiones en RS.

4. *¿En qué redes sociales su empresa está presente (tiene una o varias cuentas creadas)?*

■ Facebook, ■ Twitter, ■ Youtube, ■ LinkedIn, ■ Google +, Instagram,

■ Pinterest, Tuenti, _____

5. *¿Cuál de las redes sociales considera más importante para su empresa? (mencionar sólo una)*

■ Facebook, Twitter, Youtube, LinkedIn, Google +, Instagram,

Pinterest, Tuenti, _____

Facebook por volumen: Twitter es más la esencia, pero Facebook tiene más volumen.

Acciones más grandes en Facebook que en Tw.

6. *¿Cómo se realiza la gestión de las redes sociales? (internamente, subcontratada, parcialmente subcontratada)*

Todo interno. Puntualmente se requiere la ayuda de una agencia que varía en función de la acción.

7. *Dentro de la empresa, ¿Qué personas son las que gestionan las redes sociales?*

Dividido en 4 partes:

- Comunicación: 2 pers. Difusión de los contenidos de comunicación
- Marketing: 1 pers
- Atención al cliente: 5 pers + atención 24h soporte. Solución transversal.
- Contenidos: 2 pers

8. *¿De qué departamento depende este equipo de personas?*

Respuesta implícita en la pregunta anterior.

9. *¿Monitorizan la presencia en las redes sociales?*

Si. Canales (SAC) cuantitativos: menciones

Marketing: campañas.

10. *¿Qué herramientas de monitorización utilizan para el seguimiento de su presencia en las redes sociales?*

- BrandChats
- También se utilizó Sysomos, a través del proveedor: Lewis PR. Para seguimiento de campañas.

11. *¿Con qué frecuencia emiten informes de seguimiento de la actividad en las redes sociales?*

No un modelo a partir de una plantilla base para seguimiento de campañas con periodicidad según la campaña: diarias, semanales o mensuales.

12. *¿Qué métricas consideran como las más importantes a monitorizar?*

Tráfico, comparativas con otras campañas

Cualitativas: en desarrollo. Son más complicadas de comparar debido a las diferencias entre campañas.

13. *¿Han utilizado alguna vez, las herramientas de monitorización para identificar posibles insights?*

SAC: hacen seguimiento de lo que se dice, que sirve al hacer una nueva campaña.

Cuando se prepara una nueva campaña, se hacen focus group y junto a los informes del focus se añaden informaciones de lo que ha aparecido en las redes sociales.

Considera que los clientes no hablan mucho de los productos financieros en las redes sociales. En el sentido de opinión sobre algún producto o servicio.

Si, junto con focus, entrevistas, sólo no, porque los clientes no están sólo en las redes sociales y porque considera que falta información.

La información de las redes sociales es interesante porque puede ser muy impulsiva, pero no se tienen muchos datos de los usuarios y al ser muy impulsiva es poco reflexiva que es lo que se necesita para elaborar un producto como una hipoteca.

Información que se tiene que coger con “pinzas”

14. *¿Monitorizan la métrica “Sentiment” (valora los comentarios positivos, negativos o neutros) que realizan sus consumidores?*

Aparece en todos los informes, pero no se lo cree de ninguno.

Lleva 5 a 6 años en la gestión de las RS pero nunca ninguna herramienta ha tenido un % de acierto suficientemente elevado como para darle crédito.

No identifican la ironía, dobles sentidos, falta evolucionarlas.

15. *¿Consideras que será una variable importante en un futuro?*

Lo que evolucionará será lo que decimos, la red social en sí, la evolución de las propias plataformas.

Lo que se haga en la propia red es lo que evolucionará

Sentiment evolucionará pero en estos momentos, no le ve la utilidad clara.

Los usuarios expresan más sentimientos negativos que positivos. Se hacen más críticas fáciles y banales que reflexivas.

Mejor analizar más dentro de la globalidad y las tendencias que en concreto.

Los comentarios de Ferran Pallàs se realizan dentro del marco de las RRSS del sector bancario y no de las RRSS en general.

3.- Entrevista:

Empresa: Nestlé

Nombre: Amets Uriarte

Cargo: Digital Acceleration Team Manager

1. *¿Cuál es el papel de las redes sociales en su empresa?*

Papel crítico por lo que se refiere a la relación directa con el consumidor.

Hasta hace poco, la relación con el consumidor se hacía de una manera más unidireccional y hoy día, las redes sociales nos permiten, no sólo transmitir nuestros mensajes sino también poder entenderles, interactuar con ellos y poder captar muchos más *insights* y mucha más información de la que teníamos antes. Representa una plataforma de contacto directa y por lo tanto es vital en nuestro contacto con el consumidor.

2. *¿Por qué su empresa está presente en las redes sociales?*

Es una oportunidad para poder dialogar directamente con los consumidores, entender cual su opinión sobre los productos, como poder mejorar, incluso si nuestras comunicaciones les parecen atractivas o no. Consideramos que es una piedra angular de nuestra comunicación.

Es cierto que hasta ahora, las redes sociales no se podían considerar como Mass media, pero la verdad es que cada vez tienen una penetración mayor, como facebook que tiene 19 millones de usuarios activos en España.

Las redes sociales ya tienen un alcance nada desdeñable con lo que también se utilizan como plataforma publicitaria.

3. *¿Cuáles son los principales beneficios que espera conseguir con su presencia en las redes sociales?*

- Contacto directo con el consumidor
- Este contacto permite seguir construyendo este “Equity” que tienen las marcas como Nestlé, de una manera más cercana a los consumidores, para seguir siendo “Top of Mind”
- Obtención de *insights*: estamos entendiendo qué es lo que les gusta, lo que no, identificando nuevas oportunidades, vías de innovación, de creación de nuevos servicios,
- Ayuda a vender: de forma indirecta, por supuesto.
- Fidelización: si podemos medir el “engagement” como paso previo a la fidelización, al medir los diferentes niveles de interacción de los usuarios con la marca, como paso previo y directo de medir la fidelización. Esos consumidores son más fieles.

4. *¿En qué redes sociales su empresa está presente (tiene una o varias cuentas creadas)?*

- Facebook ■ Twitter, ■ Youtube, ■ LinkedIn, ■ Google +, ■ Instagram,
■ Pinterest, Tuenti, _____ todas y con algunas pruebas.

5. *¿Cuál de las redes sociales considera más importante para su empresa? (mencionar sólo una)*

- Facebook, Twitter, Youtube, LinkedIn, Google +, Instagram,
Pinterest, Tuenti, _____

La más importante: Facebook por “reach”. Medio publicitario

Cada vez más presentes en Twitter

Youtube: líderes en recetas.

6. *¿Cómo se realiza la gestión de las redes sociales? (internamente, subcontratada, parcialmente subcontratada)*

Labor del Community manager está internalizada, por ser un rol muy estratégico.

La propuesta creativa de contenido se sigue desarrollando en las agencias.

Gestiona a nivel de marca, no por producto.

7. *Dentro de la empresa, ¿Qué personas son las que gestionan las redes sociales?*

Community Managers.

Que están en contacto directo con los negocios de los productos. Asistiendo a las reuniones de estrategia. Coordinado a nivel de producto.

8. *¿De qué departamento depende este equipo de personas?*

Depende de Marketing

9. *¿Monitorizan la presencia en las redes sociales?*

Se monitoriza: en tiempo real de cómo está funcionando los posts y los contenidos que se están publicando en las redes sociales. Desde que publica se mide alcance, para ajustar la inversión en medios, etc,...

Por otro lado se monitoriza la identificación de *insights* para la generación de contenidos en redes sociales. *Insights* para: innovar, campañas de publicidad, generar contenido.

Cada Com Man. Combina diferentes herramientas para identificar cuál es la conversación en relación con esta marca, para identificar oportunidades, inquietudes, para generar contenido para publicar en la red social,....

10. *¿Qué herramientas de monitorización utilizan para el seguimiento de su presencia en las redes sociales?*

Radian 6 y herramientas de Google. Alexa y SocialBakers.

11. *¿Con qué frecuencia emiten informes de seguimiento de la actividad en las redes sociales?*

Habitualmente de forma semanal y mensual.

12. *¿Qué métricas consideran como las más importantes a monitorizar?*

- Depende de si se evalúa el contenido, el alcance,... y de la red social
- Tanto cuantitativas como cualitativas.
- Cuantitativas: alcance, engagement, interacción,...
- Cualitativas: saber si se está conectando con el consumidor con los *insights* que se habían identificado previamente. Se adecuar en la red social,...roductadecs *insights* que se habudes, para generar contenido para publicar en la red social,...roductúa el mensaje en función de los *insights* obtenidos y la tipología del consumidor.

13. *¿Han utilizado alguna vez, las herramientas de monitorización para identificar posibles insights?*

Si, ya se ha mencionado anteriormente

14. *¿Monitorizan la métrica “Sentiment” (valora los comentarios positivos, negativos o neutros) que realizan sus consumidores?*

Si se analiza la variable “Sentiment” , lo que sí es cierto que las herramientas que se dispone hoy, no tienen la calidad suficiente como para poder ofrecer unos resultados interesantes desde el

punto de vista de sentimientos. Tenemos que intervenir muy activamente en relación a la máquina. Sí da buenos resultados en inglés, pero no en castellano y menos en catalán.

Pero es muy importante para nosotros para poder identificar posibles crisis y en un futuro será importante. Vital ya que una empresa de gran consumo tiene que tener identificadas las conversaciones que permitan identificar posibles crisis, posibles problemáticas, donde la identificación del sentimiento es muy importante.

4.- Entrevista:

Empresa: BBVA

Nombre: Saúl Rodríguez Sánchez

Cargo: Responsable de Comunicación Online, RRSS y Crisis

1. *¿Cuál es el papel de las redes sociales en su empresa?*

Comunicación, en el sentido más amplio; atención al cliente; gestión de crisis. Desde el centro corporativo se ha diseñado una política de gestión, acción y reacción en redes sociales que sea común en todos los países, así como una serie de mensajes alineados con el posicionamiento que quiere conseguir el banco. Las redes sociales son perfectas para ello.

Nuestras redes sociales sirven al Grupo como altavoz a través del cual podemos contar, cuando y como queremos, en los formatos y la intensidad que deseamos, tantos mensajes como estimemos. Pero no solo eso, sino que también nos sirven para escuchar a nuestros stakeholders. Saber qué es lo que desean, cuáles son sus inquietudes, sus quejas, las tendencias del mercado...y anticiparnos con soluciones hechas a medida.

2. *¿Por qué su empresa está presente en las redes sociales?*

BBVA está inmerso en un proceso de transformación digital desde hace tiempo, con amplia experiencia en banca online en todos los países en los que está. Con la aparición de las redes sociales como concepto, los clientes ‘viven’ y se desarrollan en ellas. Así, es normal que las empresas estén allí. En el caso de BBVA, su intención es hacer la vida más fácil a las personas.

Nuestro punto de arranque es siempre por qué, para qué y cómo vamos a estar presentes. Cualquier perfil nuevo que abrimos parte de un estudio consensuado entre el demandante del perfil (sea en la red que sea) y un acompañamiento para hacer que la apertura y gestión

respondan a la necesidad y se ajusten a lo que el público objetivo al que nos dirigimos puede demandar.

3. *¿Cuáles son los principales beneficios que espera conseguir con su presencia en las redes sociales?*

Comunicación, en el sentido más amplio, como ya he dicho en el punto anterior; Por supuesto, también la atención al cliente así como la gestión de posibles crisis de forma proactiva.

4.: *¿En qué redes sociales su empresa está presente (tiene una o varias cuentas creadas)?*

■ Facebook, ■ Twitter, ■ Youtube, ■ LinkedIn, ■ Google +, ■ Instagram,
■ Pinterest, , Tuenti, ■ Vine _____

5. *¿Cuál de las redes sociales considera más importante para su empresa? (mencionar sólo una)*

Facebook, ■ Twitter, Youtube, LinkedIn, Google +, Instagram,
Pinterest, Tuenti, _____

Facebook por su alcance masivo pero también twitter por su capacidad de llegad, interacción y Youtube por ser nuestra apuesta clara como player de contenidos en vídeo. En el caso de LinkedIn sin duda alguna es una plataforma valiosísima para nuestro equipo de Selección. Y G+ por convertirse desde ya en nuestra herramienta de colaboración interna entre todos los empleados.

6. *¿Cómo se realiza la gestión de las redes sociales? (internamente, subcontratada, parcialmente subcontratada)*

De las dos formas. Desde el departamento de Comunicación Coporativa contamos con un equipo interno, pero luego está deslocalizado según las geografías y las áreas de negocio, y en esos

casos muchas veces está externalizado con agencias pero, siempre, coordinado desde el departamento de Comunicación Corporativa correspondiente.

7. *Dentro de la empresa, ¿Qué personas son las que gestionan las redes sociales?*

La coordinación está centralizada en Comunicación Corporativa aunque luego la gestión diaria pueda estar en manos de otras áreas de negocio, como marketing, RC, etc...

8. *¿De qué departamento depende este equipo de personas?*

Comunicación Y Marca

9.: *¿Monitorizan la presencia en las redes sociales?*

Sí

10. *¿Qué herramientas de monitorización utilizan para el seguimiento de su presencia en las redes sociales?*

Hootsuite, Alto Analytics, COGNOS (SMA IBM), Radiant6, SocialBro... pero lo fundamental, las personas. Hay un equipo de personas que están atentas a lo que ocurre. Las herramientas facilitan su labor.

11. *¿Con qué frecuencia emiten informes de seguimiento de la actividad en las redes sociales?*

Diario, cada dos días, semanal, mensual... depende de las necesidades, pero lo que si que podemos garantizar es una respuesta inmediata a cada necesidad.

12. *¿Qué métricas consideran como las más importantes a monitorizar?*

Menciones negativas vs influencers/stakeholders; también el volumen de ruido de eventos, comparados entre ellos, para saber qué supone a BBVA, o mirando, en determinados casos, a la

competencia...; Ojo, no nos olvidamos de lo positivo o neutro, volúmenes de los cuales se pueden sacar bastantes conclusiones a la hora de emitir futuros mensajes, ideas o incluso productos.

13. *¿Han utilizado alguna vez, las herramientas de monitorización para identificar posibles insights?*

Sí: para descubrir nuevos stakeholders, o para intentar hacer previsiones a futuro, encontrar nuevas temáticas o líneas de conversación que se puedan estar generando alrededor de BBVA o del sector bancario.

14. *¿Monitorizan la métrica “Sentiment” (valora los comentarios positivos, negativos o neutros) que realizan sus consumidores?*

Hace años que no damos valor a la medición automática, más allá de la que tiene... siempre merece la pena hacer una cata de mensajes y etiquetarlos exhaustivamente a mano.

5.- Entrevista:

Empresa: DANONE

Nombre : Rafel Feliu

Cargo: Director del departamento Digital & CRM

1. *¿Cuál es el papel de las redes sociales en su empresa?*

Danone empezó su presencia en las redes sociales hace 4 años y ha ido evolucionando (al inicio las RS tenían un punto de viralidad que consideramos que han ido perdiendo) y aumentando la experiencia en el uso y beneficio de las redes sociales.

Cada vez más, para destacar en las RS se tiene que invertir en medios pagados, así que las RS son un buen medio para comunicar, pero además si el mensaje es bueno, se viraliza y otra de las variables es la interactividad con el clientes.

Desde el servicio de atención al cliente, a quien, ante cualquier duda, se le ofrece una respuesta lo más rápida posible.

Y como se verá más adelante, la búsqueda de *insight* o necesidades del cliente para poder ofrecer a los clientes, cada vez, mejores productos.

2. *¿Por qué su empresa está presente en las redes sociales?*

Primero: estar por estar

Ahora: plataforma para poder interactuar y comunicarse de forma directa con el consumidor.

¿Captación y Fidelización?

Captación: Si, se puede captar, mediante incentivos. Que no tienen porque ser necesariamente regalos, pueden ser contenidos interesantes que quiera compartir.

Fidelización: las RS no deja de ser un CRM que mediante contenidos, este cliente siga siendo activo en tu entorno.

Diferenciamos entre: Lead y Hot Lead

3. *¿Cuáles son los principales beneficios que espera conseguir con su presencia en las redes sociales?*

Transmisión de mensajes

Viralización

Identificación de actitudes, sensaciones del consumidor

Posible identificación de *insights*

4. *¿En qué redes sociales su empresa está presente (tiene una o varias cuentas creadas)?*

■ Facebook, ■ Twitter, ■ Youtube, ■ LinkedIn, ■ Google +, ■ Instagram,

■ Pinterest, ■ Tuenti, _____

- Facebook y Youtube fueron las primeras
- Google + para seo
- Tuenti la están dejando porque su evolución será hacia el móvil

5. *¿Cuál de las redes sociales considera más importante para su empresa? (mencionar sólo una)*

■ Facebook, ■ Twitter, ■ Youtube, ■ LinkedIn, ■ Google +, ■ Instagram,

■ Pinterest, ■ Tuenti, _____

6. *¿Cómo se realiza la gestión de las redes sociales? (internamente, subcontratada, parcialmente subcontratada)*

Inicio con agencias, con equipos de com managers

Servicio interiorizado porque lo consideran clave

Hay una persona externa de una agencia a tiempo parcial que coordina la estrategia de todas las redes sociales y coordina el equipo de community managers

Hay un equipo de Content & Community Managers: porque hay una relación muy directa entre el contenido, la línea editorial y la relación con el consumidor.

7. Dentro de la empresa, ¿Qué personas son las que gestionan las redes sociales?

Community Managers

8. ¿De qué departamento depende este equipo de personas?

Departamento de Marketing.

Dentro del cual hay el equipo de Digital.

Que da servicio al resto de marcas y se divide en diferentes formas: Paid/own/earn media

Earn: gestión RS y monitoriza el ruido

Servicio de gestión organizado de forma transversal de Danone, que se subdivide por productos.

Aunque hay “Alimenta Sonrisas” que está a un nivel superior.

Comunicación corporativa

Marketing Research

¿SAC? Está integrado en el dpto de Mkt que se ha ido digitalizando.

9. ¿Monitorizan la presencia en las redes sociales?

Si.

10. *¿Qué herramientas de monitorización utilizan para el seguimiento de su presencia en las redes sociales?*

- Alertas de Google (inicios)
- Asomo
- Linkfluence (en actual proceso de integración con esta nueva herramienta)
- Principales características: agilidad
- Método de rastreo (algoritmo) puede ser más o menos bueno pero se valora la persona que interpreta los datos.

11. *¿Con qué frecuencia emiten informes de seguimiento de la actividad en las redes sociales?*

Informes: realizan muchos informes

- Mensual: a dirección
- Mensuales por marcas
- Diarios: alertas, o de seguimiento de las marcas más grandes

12. *¿Qué métricas consideran como las más importantes a monitorizar?*

- Sentimiento + ó – (el neutro no lo tienen en cuenta)
- Volumen: datos cuantitativos
- Tendencias: sobre temáticas /temas o subtemas mediante clusters

13. *¿Han utilizado alguna vez, las herramientas de monitorización para identificar posibles insights?*

Proceso habitual: por costumbre se siguen haciendo como siempre, institutos de investigación

RS: las han utilizado para realizar pruebas de packaging, música,

Por rapidez y coste. En 1h pueden hacer una encuesta a 1.000 personas.

A nivel de packaging: encuestas al perfil social con diferentes imágenes y los usuarios votaban con un incentivo de prueba de producto. Muestras de músicas.

A partir de comentarios de clientes/usuarios (escucha) con menor frecuencia, pero se hace para obtener *insights*.

Es más difícil que requiere un esfuerzo adicional.

14. *¿Monitorizan la métrica “Sentiment” (valora los comentarios positivos, negativos o neutros) que realizan sus consumidores?*

Danone ha tenido problemas con algunos productos que han tenido un impacto en las RS, que se han monitorizado y descubierto los sentimientos positivos y negativos. No tenemos en cuenta el sentimiento neutro porque es muy difícil de analizar.

Si le damos la credibilidad pero la auditamos con una persona detrás que monitoriza, para discernir entre las ironías,lo monitorizamos y le damos validez

¿Hay margen o recorrido en la variable sentiment?

Si, muchísimo. Creo que estamos a nivel 2 sobre 10 y a medida que vayan evolucionando las herramientas de análisis y se recojan más informaciones como el sentimiento y la opinión en las RS estas herramientas tendrán más relevancia.

6.- Entrevista:

Empresa: Catalunya Caixa

Nombre: María José Martínez Rodríguez

Cargo: Directora de Imagen, RRPP y Comunicación Online

1. *¿Cuál es el papel de las redes sociales en su empresa?*

Sin respuesta.

2. *¿Por qué su empresa está presente en las redes sociales?*

Sin respuesta.

3. *¿Cuáles son los principales beneficios que espera conseguir con su presencia en las redes sociales?*

Damos respuesta a las tres primeras preguntas en una única porque consideramos que están muy relacionadas y así no reiteramos.

En nuestra empresa las redes sociales las consideramos un importante canal a través del cual interactuamos con el cliente y usuarios no clientes, recogiendo sus impresiones para dar respuesta a sus necesidades y corregir el comportamiento con el que el usuario se sienta disconforme.

Asimismo, las redes sociales son una herramienta primordial para aportar valor a la marca, generando contenido exclusivamente online que sea del interés del usuario, a través del cual transmitamos los valores corporativos.

Las redes sociales, a diferencia de otros canales, nos aportan un alto grado de proximidad al usuario.

4. *¿En qué redes sociales su empresa está presente (tiene una o varias cuentas creadas)?*

■ Facebook, ■ Twitter, ■ Youtube, ■ LinkedIn, Google +, Instagram,
Pinterest, Tuenti, ■ SlideShare

5. *¿Cuál de las redes sociales considera más importante para su empresa? (mencionar sólo una)*

■ Facebook, Twitter, Youtube, X LinkedIn, Google +, Instagram,
Pinterest, Tuenti, _____

6. *¿Cómo se realiza la gestión de las redes sociales?*

Internamente. Únicamente hay subcontratación guiada por una persona interna en el caso de Facebook.

7. *Dentro de la empresa, ¿Qué personas son las que gestionan las redes sociales?*

Las redes son gestionadas por personas especialistas en cada uno de los segmentos a los que se dirige cada perfil.

Un responsable de equipo de Comunicación gestiona el Canal YouTube, el perfil en SlideShare, el de LinkedIn y el perfil de Twitter de novedades/actualidad.

Tres técnicos de Atención Multicanal alternan diariamente la gestión del perfil de atención al cliente.

Un responsable de proyecto de Marketing, junto con una empresa externa, gestiona el perfil de Facebook.

Un técnico de la Gestora de fondos de inversión gestiona el perfil especializado en esta temática.

8. *¿De qué departamento depende este equipo de personas?*

Cada una depende del departamento implicado en el contenido:

Imagen, RRPP y Comunicación Online

Atención multicanal

Marketing de particulares e innovación

Gestora inversión

No obstante, Imagen, RRPP y Comunicación Online realiza también la tarea de coordinar la interacción entre perfiles y de asesoramiento.

9. *¿Monitorizan la presencia en las redes sociales?*

Sí. Monitorizamos presencia en redes, medios online, foros y blogs.

10. *¿Qué herramientas de monitorización utilizan para el seguimiento de su presencia en las redes sociales?*

Combinamos herramientas gratuitas como pueden ser Hootsuite, Google Alerts con un servicio de pago que recoge datos de Radian6 para redes sociales y foros, y de ePrensa para medios online y blogs. Anteriormente también habíamos utilizado iMente.

11. *¿Con qué frecuencia emiten informes de seguimiento de la actividad en las redes sociales?*

Realizamos seguimiento diario, con sistema de alertas al instante, y estadísticas mensuales.

12. *¿Qué métricas consideran como las más importantes a monitorizar?*

El volumen de interacción con contenidos. Por ejemplo, en el caso de Twitter, los RT, FV, respuestas y menciones. Y en el de Facebook, los me gusta, comentarios y participaciones. Inicialmente dimos importancia al número de seguidores, pero tras los primeros años primamos el volumen de interacción.

13. *¿Han utilizado alguna vez, las herramientas de monitorización para identificar posibles insights?*

A partir de la monitorización, valoramos las temáticas que más agradan o detectamos temas nuevos que puedan interesar. Sí que la utilizamos para identificar contenidos que podríamos reiterar o los que debiéramos evitar.

14. *¿Monitorizan la métrica “Sentiment” (valora los comentarios positivos, negativos o neutros) que realizan sus consumidores?*

Sí, es una métrica a la cual le damos mucha importancia. En caso necesario, intentamos corregirla mediante la detección de *insights* anteriormente citada.

Tabla de los diez primeros usuarios por orden de importancia

Cod Usu.	NomUsuari	Imp. %	SentiPos	SentiNeg	Data	Oració central	Paraules Clau	Tot Adj dif	Tot. Adv dif	Tot. Conj dif	Tot. Conj. Sub.	Tot. Verbs	Total Oraci.	Long.prom. oració:	Tot pal.
83	Antonio Matarranz	0,768%	29	-46	Wed Dec 11 23:08:14 +0000 2013	La ejecución hipotecaria bajo sospecha ¿Hay Derecho? http://t.co/C3q5I6cWKN vía @HayDerecho http://hayderecho.com/2014/01/08/la-ejecucion-hipotecaria-bajo-sospecha-de-inconstitucionalidad-y-contrariedad-al-derecho-europeo/ Neymar, una respuesta para cada profecía - http://t.co/eyj7NGBDUy vía @ELSEISDOBLE http://www.e6d.es/otros-temas/neymar-una-respuesta-para-cada-profecia	risa derecho	33	8	4	3	41	53	19,06	1010
17	EL SEIS DOBLE	0,685%	71	-136	Mon Dec 16 21:25:11 +0000 2013	@RubenGlezAmoedo es lo malo de las coberturas, k a veces te salvan y otras te fastidian, jeje muy buen aporte. http://bit.ly/1FS1UZv	elseisdouble vía historia viajar viajes partido partir partidos	45	15	8	5	93	81	22,49	1822
12	Roger Vila	0,650%	186	-135	Wed Dec 11 22:02:58 +0000 2013	RT @SiPeroNo1: Aunque no lo parece, tal vez haya algún dirigente del PP que no sea un delincuente http://goo.gl/rSELNk	buenas buena bueno buenos buen risa herreratr bajar bajará bajando bajada baja bajaría parece parecido parecen parecidos malo mal mala	82	38	14	8	182	104	22,68	2359
9	Francisco Cantero	0,616%	116	-147	Sun Dec 15 10:04:48 +0000 2013	La ley del aborto se llamará "de protección de la vida del concebido" http://t.co/OkJ3IR1j9Y vía @el_pais http://sociedad.elpais.com/sociedad/2013/12/20/actualidad/1387544028_883233.html	siperono risa dicen dice pp parece mejorando mejor	47	33	13	8	171	99	25,68	2542
125	Pedro Domnguez	0,612%	25	-47	Tue Dec 10 13:19:38 +0000 2013	Si no tienen pan, que saquen de su cuenta en Suiza http://www.lavozdeg Galicia.es/noticia/galicia/2014/02/21/angel-espadas-inocente-digo-cometiera-practicas-poco-eticas/0003_201402G21P5991.htm	vía elpais minutos aborto	31	7	4	3	85	59	20,59	1215
13	Armando el pollo	0,605%	121	-116	Sat Dec 14 09:46:38 +0000 2013	IU lamenta la caída de la población activa y dice que se crea empleo "basura" http://t.co/98GZCn3dEX vía @eldiarios http://www.eldiario.es/_cc41c21	suiza suizo cuenta risa pase pasa pasando paso pasada pasar pasado pasarán	69	25	7	4	138	100	19,89	1989
35	Germn Basoa Rivas	0,600%	41	-70	Sun Dec 08 07:53:00 +0000 2013	#Murcia: Si eres emprendedor/a por creer en el PP, trabajas 10h/día y los resultados te dicen que sólo cubres gastos... No eres feliz. http://www.que.es/ultimas-noticias/espana/201210120900-hispanidad-abucheos-zapatero-conazo-raioy-cont.html	eldiarios vía ley madridcapital aborto abortar	32	10	7	5	72	59	21,27	1255
24	Ani Olano Excelsior	0,596%	142	-167	Sat Dec 14 20:37:23 +0000 2013	El PP avala que se pueda cortar en invierno la luz por impago http://t.co/VBIRT5GP0u vía @publico_es hijos de http://www.publico.es/dinero/489944/el-pp-avala-que-se-pueda-cortar-en-invierno-la-luz-por-impago	murcia pp feliz risa sólo sol dicen dice dices	69	32	13	7	155	97	27,53	2670
39	pedro #28F	0,594%	50	-133	Wed Dec 11 15:12:12 +0000 2013	Locura habitacional en Alicante con miles de personas en la calle http://t.co/g8bhAqcjVi vía @sharethis http://shar.es/Qkq3r	vía publicidad publicoes pp gobierno	38	23	11	7	110	71	23,03	1635
10	Eduardo Lizarraga	0,579%	64	-94	Fri Dec 06 10:47:53 +0000 2013		vía sharethis risa nuevo nuevos nueva hoy	39	17	11	6	119	73	21,10	1540

Tabla 18: Tabla de los diez primeros usuarios por orden de importancia.

Índice de tablas

<i>Tabla 1: Comparativa de los medios masivos en relación a los medios sociales</i>	<i>p.83</i>
<i>Tabla 2: Cuadro resumen de los métodos tradicionales de investigación</i>	<i>p.123</i>
<i>Tabla 3: Tabla resumen de las técnicas proyectivas más recientes</i>	<i>p.128</i>
<i>Tabla 4: Tabla de las emociones según Parrot</i>	<i>p. 251</i>
<i>Tabla 5: Tabla de presentación de las palabras más importantes</i>	<i>p. 321</i>
<i>Tabla 6: Tabla de topics por palabra clave</i>	<i>p. 322</i>
<i>Tabla 7: Tabla de tuits del topic 1</i>	<i>p. 323</i>
<i>Tabla 8: Tabla de tuits del topic 2</i>	<i>p. 324</i>
<i>Tabla 9: Tabla de tuits del topic 3</i>	<i>p. 324</i>
<i>Tabla 10: Tabla de tuits del topic 4</i>	<i>p. 325</i>
<i>Tabla 11: Tabla de distribución de los tuits por topics</i>	<i>p. 326</i>
<i>Tabla 12: Tabla de ordenación de los usuarios por topic 1</i>	<i>p. 330</i>
<i>Tabla 13: Tabla de ordenación de los usuarios por topic 2</i>	<i>p. 331</i>
<i>Tabla 14: Tabla de ordenación de los usuarios por topic 3</i>	<i>p. 332</i>
<i>Tabla 15: Tabla de ordenación de los usuarios por topic 4</i>	<i>p. 333</i>
<i>Tabla 16: Tabla del usuario más importante</i>	<i>p. 334</i>
<i>Tabla 17: Tabla que ordena a los usuarios con mayor carga emocional</i>	<i>p. 337</i>
<i>Tabla 18: Tabla de los diez primeros usuarios por orden de importancia</i>	<i>p. 453</i>

Índice de figuras

<i>Figura 1: Ilustración que representa el origen del primer grafo.</i>	<i>p.139</i>
<i>Figura 2: Mapa del mundo del Teorema de los cuatro colores.</i>	<i>p.142</i>
<i>Figura 3: Representación gráfica del juego de Hamilton.</i>	<i>p.143</i>
<i>Figura 4: Isómeros químicos.</i>	<i>p.144</i>
<i>Figura 5: Sociograma que representa un estudio sociométrico.</i>	<i>p.148</i>
<i>Figura 6: Ilustración de los grafos (personas).</i>	<i>p.152</i>
<i>Figura 7: Ejemplo de “Grafo Estrella”.</i>	<i>p.155</i>
<i>Figura 8: Ilustración de dos posibles situaciones de agujeros estructurales.</i>	<i>p.157</i>
<i>Figura 9: Red de 20 nodos construida según el modelo de Watts & Strogatz.</i>	<i>p.157</i>
<i>Figura 10: Bonding y bridging: son dos formas diferentes de conectividad.</i>	<i>p.159</i>
<i>Figura 11: Gráfico que ilustra la hipótesis que formuló Milgram.</i>	<i>p.160</i>
<i>Figura 12: Representación de un grafo conocido con el nombre de “House”.</i>	<i>p.162</i>
<i>Figura 13: Representación de un grafo.</i>	<i>p.162</i>
<i>Figura 14: Ejemplos de grafos.</i>	<i>p.163</i>
<i>Figura 15: Ejemplo de dos grafos: (A) no dirigido y (B) dirigido.</i>	<i>p.164</i>
<i>Figura 16: Grafo que representa varios tipos de paseos (Walks).</i>	<i>p.165</i>
<i>Figura 17: La longitud del camino de esta ilustración es = 4.</i>	<i>p.166</i>
<i>Figura 18: Distancias establecidas por “capas”.</i>	<i>p.166</i>
<i>Figura 19: Ilustración de un grafo denominado “Árbol”.</i>	<i>p.166</i>
<i>Figura 20: Ilustración de la centralidad de grado.</i>	<i>p.168</i>
<i>Figura 21: Ilustración de la Intermediación o Betweenness.</i>	<i>p.168</i>
<i>Figura 22: Ilustración de un ejemplo de centralidad cercanía o closeness.</i>	<i>p.169</i>
<i>Figura 23: Ilustración de un ejemplo de Centralidad de Vector Propio.</i>	<i>p.170</i>
<i>Figura 24: Representación de dos grafos y sus matrices de adyacencia.</i>	<i>p.171</i>
<i>Figura 25: Representaciones numéricas de matrices de adyacencias.</i>	<i>p.171</i>
<i>Figura 26: Estructura de Internet al nivel de sistemas autónomos.</i>	<i>p.175</i>
<i>Figura 27: Ejemplo de red de rutas aéreas.</i>	<i>p.176</i>
<i>Figura 28: Ejemplo de mapa del metro de Washington.</i>	<i>p.176</i>
<i>Figura 29: GPS, ejemplo de la aplicación de la Teoría de Grafos.</i>	<i>p.177</i>
<i>Figura 30: Ejemplo de una red de amistades en Facebook.</i>	<i>p.178</i>
<i>Figura 31: Linaje del Análisis de Redes Sociales ARS.</i>	<i>p. 179</i>
<i>Figura 32: Resultado de una búsqueda en Google.</i>	<i>p. 182</i>
<i>Figura 33: Representación gráfica de una red de páginas de un sitio web.</i>	<i>p. 183</i>
<i>Figura 34: Red que representa la propagación de un brote de tuberculosis.</i>	<i>p. 185</i>
<i>Figura 35: Cronograma de la evolución del Resumen Automático de Textos.</i>	<i>p.194</i>
<i>Figura 36: Esquema de un proceso de abstracción de textos.</i>	<i>p. 198</i>
<i>Figura 37: Principales procesos del Resumen Automático de Textos.</i>	<i>p. 202</i>
<i>Figura 38: Principales técnicas de evaluación del RAT.</i>	<i>p. 210</i>
<i>Figura 39: El espacio ROC.</i>	<i>p. 213</i>
<i>Figura 40: Grafo basado en relaciones semánticas y de semejanza.</i>	<i>p. 223</i>

<i>Figura 41: Proceso simplificado de RAT.</i>	<i>p. 224</i>
<i>Figura 42: Captura del servicio de selección de noticias Newsblaster.</i>	<i>P. 225</i>
<i>Figura 43: Resultado de una búsqueda en Google.</i>	<i>p. 226</i>
<i>Figura 44: Captura de Twitpic (Twitter).</i>	<i>p. 233</i>
<i>Figura 45: La rueda de las emociones de Plutchik.</i>	<i>p. 250</i>
<i>Figura 46: Relaciones semánticas en WordNet.</i>	<i>p. 257</i>
<i>Figura 47: Visión en 3D del modelo Hourglass of Emotions.</i>	<i>p. 259</i>
<i>Figura 48: Gráfico de las redes sociales más utilizadas en España.</i>	<i>p. 301</i>
<i>Figura 49: Medición del “Sentimiento” en función de la hora del día.</i>	<i>p. 312</i>
<i>Figura 50: Medición del “Sentimiento” en función del día de la semana.</i>	<i>p. 314</i>
<i>Figura 51: Medición del “Sentimiento” en función del mes del año.</i>	<i>p. 315</i>
<i>Figura 52: Ranking de los Bancos Españoles por activos.</i>	<i>p. 351</i>



Esta Tesis Doctoral ha sido defendida el día ____ d _____ de 201__

En el Centro _____

de la Universidad Ramon Llull, ante el Tribunal formado por los Doctores y Doctoras

abajo firmantes, habiendo obtenido la calificación:

Presidente/a

Vocal

Vocal *

Vocal *

Secretario/a

Doctorando/a

(*): Sólo en el caso de tener un tribunal de 5 miembros



 **Blanquerna**

Universitat Ramon Llull

Facultat de Comunicació
i Relacions Internacionals Blanquerna
Universitat Ramon Llull

Grup de Recerca en Estratègia
i Creativitat en Publicitat
i Relacions Públiques

Doctorat en Comunicació