



Información, cultura y sociedad: revista del Instituto de Investigaciones Bibliotecológicas

ISSN: 1514-8327

ISSN: 1851-1740

inibi@filo.uba.ar

Universidad de Buenos Aires

Argentina

Arce García, Sergio; Menéndez Menéndez, María Isabel
Aplicaciones de la estadística al framing y la minería de texto en estudios de comunicación
Información, cultura y sociedad: revista del Instituto de Investigaciones
Bibliotecológicas, núm. 39, 2018, Diciembre, pp. 61-70
Universidad de Buenos Aires
Argentina

Disponible en: <https://www.redalyc.org/articulo.oa?id=263057241004>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UAEH
redalyc.org

Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

Aplicaciones de la estadística al *framing* y la minería de texto en estudios de comunicación



Sergio Arce García¹ y María Isabel Menéndez Menéndez²

¹Universidad Internacional de La Rioja (UNIR). Escuela Superior de Ingeniería y Tecnología (ESIT), España | sergio.arce@unir.net / <https://orcid.org/0000-0003-0578-9787>

²Universidad de Burgos, Facultad de Humanidades y Comunicación, España | mimenendez@ubu.es / <https://orcid.org/0000-0001-7373-6885>

Resumen

Las técnicas de análisis del discurso de los medios de comunicación han experimentado una gran evolución gracias a la teoría del *framing*, pero con el tiempo se ha percibido que el estudio realizado por los investigadores podría presentar una elevada subjetividad. Para solventar este problema, se han desarrollado metodologías de trabajo más objetivas, adaptando técnicas estadísticas para el examen de los marcos o encuadres principales, siendo el análisis de *cluster* uno de los elementos más significativos. Otras metodologías posteriores, como la minería de texto, llegan a plantear un análisis enteramente realizado a través de algoritmos informáticos, con conocimientos morfológicos de los distintos idiomas. En el presente texto se expone el estado de la cuestión de estas metodologías, sus herramientas más destacadas y los programas informáticos que ayudan en el análisis estadístico del *framing* con el objetivo de sistematizar las opciones actualmente disponibles para las investigaciones en comunicación.

Palabras clave

Minería de texto
Framing
Estadística
Comunicación
Programa informático

Abstract

Application of statistics to framing and text mining in communication studies. Techniques of discourse analysis in the media have undergone a great evolution thanks to the 'framing' theory, but over time it has been perceived that the researchers' studies could include high levels of subjectivity. To solve this problem, more objective methodologies have been developed, adapting statistical techniques to the examination of the main frames; cluster analysis being one of the most significant elements within these processes. More recent methods, such as text mining, propose the analysis to be entirely done through computer algorithms with morphological knowledge of different languages. This article approaches the state of the art of these methodologies, their most outstanding tools and the computer softwares that help in the statistical

Keywords

Text mining
Framing
Statistics
Communication
Software

analysis of framing, with the aim of systematizing the options that are currently available for communication research.



Artículo recibido: 11-04-2018. Aceptado: 20-07-2018

1. Introducción

Con la evolución de las teorías comunicativas, y de su aplicación al análisis empírico del discurso de los medios de comunicación, durante los últimos años hemos asistido a la aparición de numerosos enfoques académicos que, esencialmente, han desarrollado la técnica del *framing* en España (Ardèvol Abreu, 2015). Sin embargo, el uso de esta técnica se suele circunscribir a la búsqueda manual de los encuadres o marcos principales, en su mayor parte de determinadas palabras, a los cuales se asocian una serie de propiedades que posteriormente son contabilizadas mediante algún programa informático estadístico para determinación de porcentajes y variaciones temporales. Actualmente estas técnicas de análisis avanzadas se utilizan en distintas ramas del ámbito científico en análisis de textos y son por tanto plenamente válidas para su uso común y de análisis de los medios de comunicación y las redes sociales.

Dado que la bibliografía en español es todavía escasa, en este artículo se ofrece una revisión e interpretación del estado de la cuestión existente en el ámbito científico mundial sobre el tratamiento de textos y su análisis de marcos clave, aunque con mayor interés por el espacio de habla hispanohablante, con la pretensión de ofrecer una síntesis de posibilidades metodológicas aplicables al campo de la comunicación. El objetivo general del texto se fundamenta en que estas técnicas son relativamente recientes y su uso es más elevado en países anglosajones y asiáticos por lo que el interés que guía el artículo parte de la posibilidad de aplicación en el diseño de futuros trabajos, al ofrecer a los investigadores un recopilatorio de las metodologías disponibles en la actualidad, sus ventajas e inconvenientes, su evolución en el tiempo y las posibilidades de utilización.

El empleo aún muy inicial de estas técnicas explica que la literatura académica sea todavía escasa, además de encontrarse en continua evolución, principalmente en áreas de estudio informático para análisis masivo de textos, traducciones automáticas y determinación de patrones. Dichos estudios se encuentran más avanzados principalmente en lengua inglesa y japonesa, pero existen investigaciones iniciales en otras lenguas latinas, incluido el castellano.

A partir del objetivo general descrito en las líneas precedentes, se establecieron los siguientes objetivos secundarios:

- » Identificación de las técnicas estadísticas que pueden emplearse en el análisis de textos en el ámbito de la comunicación que introducen posibilidades de neutralizar la subjetividad del investigador.
- » Definición de las distintas formas de aproximación a la búsqueda de marcos principales, desde la identificación realizada por humanos a la automatizada mediante algoritmos matemáticos.
- » Descripción de los programas informáticos que permiten desarrollar cada una de las vías anteriormente descritas.
- » Análisis de las posibilidades futuras de estas técnicas en la investigación en comunicación.

Mediante los objetivos planteados, se pretende aportar una visión en conjunto de lo que las nuevas tecnologías pueden ofrecer al análisis de la comunicación, permitiendo investigaciones lo más objetivas posibles además de la utilización de corpus de investigación de gran tamaño, posibilidad muy atractiva tanto para las investigaciones sobre prensa y otros medios de comunicación como las interesadas por redes sociales.

2. Metodología

Para el cumplimiento de los objetivos marcados, se ha realizado una revisión exploratoria del estado de la cuestión en las principales bases de datos de artículos indexados, tanto en lengua española como inglesa, aunque incidiendo principalmente en el uso de estas aplicaciones en castellano. Dentro de este proceso de revisión se han buscado las expresiones “text mining”, “minería de texto” y “framing”, discriminando como criterio de exclusión todos aquellos resultados que no estén aplicando soluciones estadísticas a los análisis y conclusiones obtenidas. Las bases de datos empleadas para la búsqueda han sido Web of Science, Scopus y Dialnet.

Como la pretensión de este artículo no es hacer una mera recopilación sobre el estado de la cuestión sino sintetizar las posibles metodologías y aplicaciones estadísticas sobre el texto comunicativo, aunque en la revisión exploratoria se llevó a cabo una investigación de tipo sistemático y de metaanálisis, los porcentajes de inclusión y exclusión no se han considerado relevantes. No obstante, la bibliografía que cumple las condiciones establecidas es reciente, escasa y proviene principalmente del campo informático, siendo casi inexistente en el campo social y comunicativo de habla hispana, e incipiente en lengua inglesa o japonesa.

Por otro lado, se han buscado en Internet las aplicaciones informáticas existentes que permiten el análisis estadístico de textos comunicativos en las técnicas de *framing*, sea por identificación manual de los marcos principales por parte del investigador o bien mediante un algoritmo matemático por técnica de minería de texto, empleándolas para comprobar sus posibilidades en la investigación de los medios de comunicación y las redes sociales.

3. ¿Qué es el *framing*?

A lo largo de las décadas de los sesenta y setenta del siglo XX, la investigación académica estuvo dominada por una corriente objetiva que pretendía mostrar la información al público sin ningún tipo de distorsión. Ya en los ochenta, aparece un nuevo enfoque comunicativo: la teoría del *framing*, cuya hipótesis de partida es que, en los medios, cada periodista interpreta la realidad a partir de filtros, conscientes e inconscientes que, por su misma subjetividad, impiden que “un mismo hecho sea interpretado y contado de la misma forma por dos personas” (Giménez Armentia y Berganza Conde, 2009).

El concepto del *framing* o encuadre noticioso proviene de Erving Goffman (1974), que lo adapta a los medios de comunicación desde la sociología. La idea principal es que cada persona enmarca una situación mediante esquemas individuales y sociales, siendo esta última la principal. Así, los medios no pueden entenderse como meros transmisores de información, sino que seleccionan hechos de la realidad para presentarlos en un contexto social y cultural, asumiendo un impacto sobre la lógica y el conocimiento del público (Sádaba Garraza, 2001). El *framing* parte de una premisa: el tratamiento del acontecimiento y de las demás inserciones periodísticas van a producir

un cambio en las opiniones y actitudes del receptor del mensaje “debido a sutiles alteraciones en la definición del problema” (Igartúa y Humanes, 2004).

La metodología del *framing* sigue un patrón que según López Rabadán (2010) consta de cuatro fases: la primera consiste en la revisión y selección de los textos para la construcción de la muestra correcta y su representatividad. Para ello, se deben identificar las conexiones culturales, ideológicas, periodísticas, encuadres y proceso mediático. La segunda fase consiste en la realización de un análisis estadístico de la agenda temática que permite delimitar las frecuencias más representativas. A partir de estas frecuencias se realiza una descripción y contextualización de los términos. La tercera fase tiene en cuenta las palabras clave como representación discursiva del proceso del *framing* y hace un estudio interpretativo para identificar presencias y ausencias. Finalmente, la cuarta y última fase es la propuesta de conclusiones generales sobre el encuadre, de manera que se verifique la estrategia tomada y la identificación crítica de las tendencias recogidas en el análisis.

El *framing* se inició con el estudio manual de identificación de las palabras o marcos clave y su relación respecto a otros términos o encuadres encontrados. La estrategia se basa principalmente en el conteo y estudio subjetivo de dichas relaciones. Para evitar estas subjetividades a la hora de categorizar los encuadres, Miller y Riechert (2001) propusieron la aplicación de programas informáticos que permitieron identificar los marcos. Dichos autores nombraron la técnica como *frame mapping*. En ella se combina el análisis computarizado a escala multidimensional, escogiendo inicialmente determinadas palabras y frases que se estudian estadísticamente a través de análisis de *cluster* y de co-ocurrencias.

4. La minería de texto

Esta técnica informática y estadística comenzó a ser expuesta en los textos académicos a principios de los años noventa y ha ido evolucionando con el paso de los años. Se trata de una metodología que permite extraer conocimiento a partir de grandes cantidades de datos y textos, mediante la búsqueda de patrones y referencias, encontrando así información no explícita con la mínima intervención o interferencia humana (Lin, Hao, Liao, 2016). Para Justicia-de-la-Torre, la minería de texto se puede definir como “El proceso que descubre información útil que no está presente explícitamente en ninguno de los documentos objeto de análisis y que surge cuando se estudian adecuadamente y se relacionan dichos documentos” (2017: 28). Según esta misma autora, la metodología es útil en multitud de aplicaciones en la sociedad de la información y puede ser usada para resumir contenidos y sintetizar información, para la extracción de ideas útiles y el análisis de opinión, para redes sociales, marketing y comercio electrónico, así como para estudiar la relación de Internet con sus usuarios.

La principal diferencia de la minería de texto como método respecto al *framing* proviene de quien toma los datos de muestra y establece los marcos. En la minería de texto es el sistema informático, a través de una serie de técnicas estadísticas y algoritmos matemáticos, quien realiza la identificación y el primer análisis de los datos que posteriormente deberán ser verificados por el investigador. Es decir, se realiza una organización y clasificación puramente automáticas y, de la misma forma, la recuperación, extracción, evaluación y predicción de información. Dado que la intervención humana queda reducida al mínimo, se puede aseverar que se logran reducir “los problemas generados por la codificación humana de la información textual” (Álvarez-Gálvez *et al.*, 2014).

La fiabilidad y precisión de la minería de datos para su uso en textos en lengua castellana han sido corroboradas en diversos estudios, como el de Álvarez-Gálvez et al., (2014) y el de Contreras Barreras (2016). También se ha corroborado en otros idiomas (Aureli, 2017). El uso de un algoritmo matemático para la identificación de marcos favorece la reproducción y hace que se eliminen los sesgos de tipo personal del investigador. No obstante, todos los autores anteriormente citados destacan que los resultados obtenidos mediante la minería de texto deben ser analizados posteriormente y cotejados por parte del investigador para eliminar los contenidos que no tengan sentido, evitar malentendidos o situaciones no realistas. Yana y Minnhagen (2018) exponen en sus estudios que el empleo de esta técnica no presenta diferencias entre idiomas, sino que los problemas de su uso provienen principalmente del empleo de palabras polisémicas, por lo que requiere de un análisis posterior de los resultados.

La aplicación de esta metodología al estudio del área de la comunicación es reciente y muy incipiente en español. También se está comenzando a emplear en áreas como la medicina para el estudio de literatura científica (Gutiérrez Sacristán *et al.*, 2017), para la clasificación de material bibliográfico (Contreras Barrera, 2016), en la informática sobre uso y aplicación de algoritmos matemáticos de aplicación a textos (Montes y Gómez, Gelbukh, López-López, 2005; Justicia de la Torre, 2017) y en el estudio de caso en el área de comunicación (Álvarez Gálvez *et al.*, 2014; Fenoll y Rodríguez-Ballesteros, 2017).

La minería de texto, para lenguas europeas, presenta dos tipos de funcionamiento de algoritmos para la recogida y tratamiento de los textos:

- *Stemming* (raíz léxica: proveniente): técnica que se basa en el recorte de las palabras para buscar sus variantes morfológicas y agruparlas. El algoritmo debe tener cierto conocimiento del idioma para la extracción de prefijos y sufijos y quedarse con la raíz. Esta técnica es la más utilizada en los programas informáticos existentes para lenguas latinas, pero presenta, en algunas ocasiones, dos tipos de errores: falsos positivos y negativos debidos a la existencia de términos polisémicos o palabras con parecida morfología, pero de distinto significado (Hajeer *et al.*, 2017).

- *Lemmatization* (lematización): técnica que se apoya en la agrupación por morfemas. El algoritmo del sistema de análisis cuenta con un diccionario del idioma e información de la morfología para su asociación (Singh y Gupta, 2017). Esta técnica necesita un conocimiento previo de la lengua para la agrupación de palabras, por lo que podría partir con ventaja con respecto a la anterior, pero según estudios como los de Kettunen, Kunttu y Järvelin (2005), o los de Balakrishnan, Humaidi y Lloyd-Yemoh (2016) los resultados realizados con ambas metodologías y en distintos idiomas han ofrecido diferencias insignificantes.

5. Aplicación estadística al estudio de la comunicación

Siguiendo las etapas descritas por López Rabadán (2010) y que se han detallado dos capítulos atrás, el análisis estadístico que se realiza en la segunda fase del *framing*, se suele reducir a un conteo del número de las palabras inicialmente seleccionadas por el investigador. Para poder realizar un estudio más exhaustivo se pueden emplear herramientas estadísticas que permiten analizar mayor cantidad de datos y de relaciones entre sí, apreciando relaciones que no son evidentes a priori pero que podrían ser útiles una vez detectadas (Williamson, Feyer y Cairns, 1996). Las técnicas que pueden ofrecer información para el análisis son las siguientes:

- » Frecuencia de códigos o marcos a través del número de apariciones que un marco o palabra aparece en el texto. Se trata del análisis más sencillo, que puede ser complementado con el conteo de veces que un marco se relaciona con otro o aporta un valor. De esta forma, se consideran los aspectos positivos, neutrales o negativos sobre el marco analizado. De momento, en los estudios de comunicación de *framing* en el ámbito español y latinoamericano este proceso se realiza de forma manual, pero en los sistemas informatizados puede realizarse a través del uso de palabras que se prefijen con atributos. Así se podrá determinar cómo se presentan los marcos no solo de forma global, sino que, al tratar grandes cantidades de informaciones a lo largo de los días, meses o incluso años, se podrán analizar evoluciones temporales en dichas atribuciones o percepciones.
- » Análisis de conglomerado o de *cluster*, técnica de análisis de datos para la clasificación en grupos ordenados que presentan entre ellos un grado de similitud. Es una de las metodologías que mejores resultados ofrece en la identificación de textos, marcos y su cobertura en el uso de *framing* (Matthes y Kohring, 2008). Presenta dos tipos de análisis, el jerárquico y el de K-medias, siendo más empleado el primero ya que no es necesario conocer de forma previa el número de agrupaciones existentes (Legara, Monterola y David, 2013; Baojun, Hua y Ye, 2017). Los análisis jerárquicos para textos se realizan principalmente a través de algoritmos por el método de Ward (1963). El análisis de *cluster* permite conocer, a través de la estadística, los discursos narrativos que los textos ofrecen. El análisis puede ser realizado por palabras o por textos completos, por lo que en su estudio se deberá seleccionar si el análisis de clasificación debe establecerse para discurso (palabras) o inserciones (documentos).
- » Análisis de correspondencias, técnica que analiza tablas de contingencia para construir diagramas cartesianos en dos o tres dimensiones entre las variables, de forma que en el gráfico resultante la proximidad entre puntos establece el nivel de asociación (Castrillo Castrillo, 2014). En el ámbito de la comunicación esta técnica puede ser empleada entre distintas inserciones periodísticas para comprobar si se establece un discurso mantenido en el tiempo o existe dispersión y falta de continuación.
- » Análisis de co-ocurrencias, permite explorar las palabras que presentan un patrón de aparición semejante. De esta forma se puede apreciar la asociación entre sí por parte de las más empleadas y los términos que las rodean (Danowski, 1993). Con este análisis se pueden observar las relaciones entre palabras y sus inserciones periodísticas (Higuchi, 2016). Para comprobar la relación existente entre los distintos términos, se analiza cada palabra con las cinco que le anteceden y las cinco posteriores; los programas informáticos buscan entonces la frecuencia de aparición y su relación a través de la definición de una variable, según lo establecido por Scott (2001), que creó un programa informático denominado *Wordsmith Tools*. Entre las distintas variables de asociación que propone, el más empleado es el logaritmo de *likelihood* (probabilidad, posibilidad) para la identificación de frases significativas que asocia a frecuencias, números de palabras relacionadas, anchura de términos y posiciones.
- » Distribución de relación logarítmica entre aparición de términos global respecto a términos por documento. Con este estudio se puede comprobar si la distribución de las palabras se produce de manera uniforme o en pocos documentos. Este análisis servirá para acotar el número de términos más destacables que deberán ser analizados posteriormente. Si se busca una palabra en concreto, se puede ver su dispersión y uso a lo largo de los distintos documentos analizados, apreciando si solo aparece en unos pocos o no.

El empleo de cualquiera de las técnicas estadísticas descritas permite afrontar el análisis en la metodología de *framing*, revelando las diferencias entre los distintos

marcos o encuadres. La diferencia principal proviene de la forma de pre-procesado e identificación de la fuente u origen de los datos, ya se trate de una detección humana o de una informatizada y automática.

6. Programas informáticos de análisis

La determinación de encuadres puede realizarse mediante programas informáticos cuya forma de entrada estará determinada en función de si es una ayuda a la identificación de marcos por parte del investigador o si es el sistema informático el que realiza una minería de texto. Las distintas posibilidades de estadística son muy variadas, aunque existe una gran disparidad de uso en la estadística aplicada. Atendiendo a quién realiza la identificación de marcos existen, de manera no exhaustiva:

- Programas de identificación manual de marcos con ayuda informática y tratamiento posterior: Atlas.ti, NVivo, QDA Miner, MaxQDA y Dedoose. Estos programas son comerciales, aunque el programa QDA Miner tiene una versión gratuita denominada QDA Miner Lite que permite hacer el análisis inicial de identificación y conteos. Presentan en su mayoría gestión estadística, aunque algunos, como Atlas.ti, requieren pasar los datos a otros programas informáticos externos específicos de estadística.

- Programas basados en identificación informática de minería de textos: Wordstat (que puede ir ligado y ser ejecutado junto al programa QDA Miner) y Sas Text Analytics son comerciales, pero existen programas de software libre entre los que cabe citar Carrotsz, Wordsmith Tools y los basados en el software estadístico R: R.temis y KH Coder. Wordsmith Tools está pensado principalmente para análisis de co-ocurrencias en inglés, mientras que Carrotsz está pensado principalmente para análisis en páginas webs, pero los basados en R presentan estructuras que incluyen el análisis de lengua castellana y estadística completa. Igualmente permiten el análisis de lo que se denomina como emociones a través del seguimiento del uso de determinadas palabras a lo largo del tiempo en los textos analizados.

7. Conclusiones

El análisis por *framing* de distintos textos de comunicación presenta actualmente avances significativos respecto a los estudios realizados en los inicios de la teoría. En el origen, se verificaba un fuerte componente subjetivo por parte del investigador en las investigaciones que se centraban en el estudio de unas pocas palabras. Con el tiempo, se han ido añadiendo técnicas estadísticas que han ido aportando un componente menos subjetivo y que han hecho aparecer nuevas conexiones entre distintos marcos que los investigadores no podrían ver o determinar claramente.

Aunque inicialmente la mayoría de las investigaciones se centraban en el conteo y atribuciones subjetivas, actualmente se pueden emplear otras técnicas de uso estadístico que devuelven mayor información. Entre ellas, es el análisis de *cluster* o agrupaciones el más empleado y el que estructura y aporta nuevas referencias en el análisis, revelándose como una forma más objetiva y menos sujeta al sesgo del investigador en la determinación de relaciones y encuadres. Otras técnicas estadísticas ayudan igualmente en la determinación del estudio de marcos principales y sus relaciones con otros secundarios; además permiten observar su evolución temporal y también la asociación de apreciaciones positivas, negativas o neutras debidas al mayor o menor uso de determinadas palabras que dan cierta connotación al texto. Entre ellas, se pueden citar el análisis de co-ocurrencias, de correspondencias y de distribución logarítmica de un término entre un documento y la totalidad de ellos. Aunque todos

los resultados deben ser verificados por el investigador, ya que algún resultado puede no tener sentido o no ser relevante, ofrecen una visión más objetiva y complementaria del análisis de *framing*.

La diferencia actual entre las técnicas de análisis se divide entre la identificación de los marcos por parte de una persona o por parte de algoritmos matemáticos a través de la metodología de minería de texto. La estrategia de pre-procesado de los textos es lo que suele determinar el posterior análisis y la necesidad de revisar los resultados. La minería de texto ofrece la posibilidad de trabajar con multitud de textos distintos de manera automatizada de forma mucho más rápida, mediante un análisis exhaustivo y objetivo, sin intervención humana durante el proceso de identificación y análisis. Por el contrario, presenta problemas ante la existencia de palabras polisémicas, términos con la misma raíz morfológica o la utilización de recursos retóricos, como pueden ser la ironía o el humor. De ahí que sea imprescindible la revisión final para el posible descarte de resultados no coherentes o fiables.

Actualmente existen distintos programas informáticos para la identificación, análisis y posterior estudio estadístico. No todos presentan las mismas características y unos se centran en un pre-procesado humano y otros en la minería de texto. Asimismo, algunos de ellos necesitan un programa anexo de análisis estadístico de los datos obtenidos porque no lo contienen.

Con la utilización e imbricación de estas técnicas, se abre una sugerente posibilidad del análisis en continuo de grandes estructuras de texto en comunicación, lo que permite estudios en profundidad sobre temáticas diversas. En estas investigaciones será factible observar sus variaciones así como las distintas relaciones entre medios y a lo largo del marco temporal.

Referencias bibliográficas

- » Álvarez Gálvez, Javier; Juan F. Plaza; Juan-Antonio Muñiz y Javier Lozano Delmar. 2014. Aplicación de técnicas de minería de textos al frame analysis: identificando el encuadre textual de la inmigración en la prensa. En *Estudios sobre el Mensaje Periodístico*. Vol. 2, 919-932.
- » Ardèvol Abreu, Alberto Isaac. 2015. Framing o teoría del encuadre en comunicación. Orígenes, desarrollo y panorama actual en España. En *Revista Latina de Comunicación Social*. Vol. 70, no. 4, 423-450.
- » Aureli, Selena. 2017. A comparison of content analysis usage and text mining in CSR corporate disclosure. En *The International Journal of Digital Accounting Research*. Vol. 17, 1-32.
- » Balakrishnan, Vimala; Norshima Humaidi y Ethel Lloyd-Yemoh. 2016. Improving document relevancy using integrated language modeling techniques. En *Malaysian Journal of Computer Science*. No. 1, 45-55.
- » Baojun, Ma; Yuan Hua y Wu Ye. 2017. Exploring performance of clustering methods on document sentiment analysis. En *Journal of Information Science*. Vol. 31, no. 1, 54-74.
- » Carrillo Castrillo, Jesús Antonio. 2014. Caracterización de la accidentalidad laboral en el sector industrial andaluz en el período 2003-2008: Aplicaciones en el diseño y evaluación de programas de intervención. Departamento de Organización Empresarial y Gestión de Empresas II (Grupo de Ingeniería de Organización). Universidad de Sevilla, España. Tesis de doctorado.
- » Contreras Barrera, Marcial. 2016. Minería de texto en la clasificación de documentos digitales. En *Biblios: Journal of Librarianship and Information Science*. Vol. 64, 33-43.
- » Danowski, James. 1993. Network Analysis of Message Content. En Richards Jr., Barnett. *Progress in communication sciences IV, USA*: Ablex. p. 197-221.
- » Fenoll, Vicente y Paula Rodríguez Ballesteros. 2017. Análisis automatizado de encuadres mediáticos. Cobertura en prensa del debate 7D 2015: el debate decisivo. En *El Profesional de la Información*. Vol. 26, no. 4.
- » Giménez Armentia, Pilar y María Rosa Berganza Conde. 2009. *Género y medios de comunicación: un análisis desde la objetividad y la teoría del framing*. Madrid: Fragua.
- » Goffman, Erving. 1974. *Frame analysis: An essay on the organization of experience*. Cambridge, USA: Harvard University Press.
- » Gutiérrez Sacristán, Alba, et al. 2017. Text mining and expert curation to develop a database on psychiatric diseases and their genes. En *Database*. Vol. 2017, 1-9.
- » Hajeer, Safaa-I.; Rasha-M Ismail; Nagwa-L. Badr y Mohamed-Fahmy Tolba. 2017. A New Stemming Algorithm for Efficient Information Retrieval Systems and Web Search Engines. En: Aboul Ella Hassanien; Mohamed Mostafa Fouad; Azizah Abdul Manaf; Mazdak Zamani; Rabiah Ahmad; Janusz Kacprzyk, eds. *Multimedia Forensics and Security*. Cham: Springer. p. 117-135.
- » Higuchi, Kiochi. 2016. KH Coder 3 Reference Manual. <http://khc.sourceforge.net/en/manual_en_v3.pdf> [Consulta: 20 marzo 2018].

- » Igartúa, Juan José y María Luisa Humanes. 2004. Imágenes de Latinoamérica en la prensa española. Una aproximación empírica desde la Teoría del Encuadre. En *Comunicación y Sociedad*. Vol. 17, no. 1, 47-75.
- » Justicia de la Torre, María del Consuelo. 2017. Nuevas técnicas de minería de textos: Aplicaciones. E.T.S. de Ingeniería Informática y de Telecomunicación. Granada: Universidad de Granada, España. Tesis de doctorado.
- » Kettunen, Kimmo; Toumas Kunttu y Kalervo Järvelin. 2005. To stem or lemmatize a highly inflectional language in a probabilistic IR environment? En *Journal of Documentation*. Vol. 61, no. 4, 476-496.
- » Legara, Erika Fille; Christopher Monterola y Clarissa David. 2013. Complex network tools in building expert systems that perform framing analysis. En *Expert Systems with Applications*. Vol. 40, no. 11, 4600-4608.
- » Lin, Fu-Ren; De Hao y Dachi Liao. 2016. Automatic content analysis of media framing by text mining techniques. Trabajo presentado al 49th International Conference on System Sciences (HICSS), celebrado en Hawaii del 6 al 7 de enero de 2016. <https://doi.org/10.1109/hicss.2016.348>.
- » López Rabadán, Pablo. 2010. Nuevas vías para el estudio del framing periodístico. La noción de estrategia del encuadre. En *Estudios sobre el Mensaje Periodístico*. Vol. 16, 235-258.
- » Matthes, Jörg y Matthias Kohring. 2008. The Content analysis of media frames: toward improving reliability and validity. En *Journal of Communication*. Vol. 58, 258-279.
- » Miller, Mark y Bonnie-Parnell Riechert. 2001. Frame mapping: a quantitative method for investigating issues in the public sphere. En West, Mark, ed. *Theory, Method, and Practice in Computer Analysis*. Westport, USA: Ablex Publishing. p. 61-76.
- » Montes-y-Gómez, Manuel; Alexander Gelbukh y Aurelio López-López. 2005. Minería de Texto empleando la Semejanza entre Estructuras Semánticas. En *Computación y Sistemas*. Vol. 9, no. 1, 063-081.
- » Sádaba Garraza, Teresa. 2001. Origen, aplicación y límites de la 'teoría del encuadre' (framing) en comunicación. En *Comunicación y Sociedad*. Vol. 14, no. 2, 143-175.
- » Scott, Mike. 2001. Comparing corpora and identifying key words, collocations, and frequency distributions through the Wordsmith Tools suite of computer programs. En Ghadessy, Mohsen; Alex Henry; Robert-L Roseberry, eds. *Small Corpus Studies and ELT: theory and practice*. Amsterdam: John Benjamins. p. 47-67.
- » Singh, Jasmeet y Vishal Gupta. 2017. A systematic review of text stemming techniques. En *Artificial Intelligence Review*. Vol. 48, no. 2, 157-217.
- » Yana, Xiaoyong y Petter Minnhagen. 2018. The dependence of frequency distributions on multiple meanings of words, codes and signs. En *Physica A: Statistical Mechanics and its Applications*. Vol. 490, 554-564.
- » Ward, Joe-H. 1963. Hierarchical grouping to optimize an objective function. En *Journal of the American Statistical Association*. Vol. 58, 236-244.
- » Williamson, Ann M.; Anne-Marie Feyer y David-R. Cairns, 1996. Industry differences in accident causation. En *Safety Science*. Vol. 24, no. 1, 1-12.