

iⁿNGENIO 2019

CONFERENCIA INTERNACIONAL DE INGENIERÍA

Agosto 21 - 23
Medellín - Antioquia

Medellín, mayo 28 de 2019

Estimados autores

Víctor Rojo - Universidad Tecnológica Nacional

Paola Britos - Universidad Nacional de Río Negro

María Florencia Pollo-Cattaneo - Universidad Tecnológica Nacional

Cordial saludo,

Les informamos que su propuesta "*Revisión de enfoques y comparación de recursos para el análisis de sentimientos en español en Twitter*", fue aprobada como ponencia oral en el Congreso IⁿNGENIO 2019.

Para que su aporte sea publicado como capítulo en libro resultado de investigación, uno de los autores se debe registrar y cancelar la inscripción como *autor*, de acuerdo con los valores en la [página de la conferencia](#).

Reiteramos el agradecimiento por confiar en nuestro evento para presentar su trabajo y esperamos contar con nuevos aportes en el futuro. Igualmente, les recordamos la responsabilidad que ustedes asumen luego de la publicación, en el sentido ético, de divulgación y de derechos de autor.

Cordial saludo,



Prof. Edgar Serna M.
Director IⁿNGENIO



Editorial
IAI

Prof. Edgar Serna M. (Ed.)

DESARROLLO E INNOVACIÓN EN INGENIERÍA

Cuarta Edición

Medellín - Antioquia
2019

Prof. **Edgar Serna M.** (Ed.)

Desarrollo e Innovación en Ingeniería

Cuarta Edición

ISBN: 978-958-52333-0-0

© 2019 Editorial Instituto Antioqueño de Investigación



Desarrollo e innovación en ingeniería [recurso electrónico] / Edgar Serna M., ed. -- 4a. ed. --
Medellín: Instituto Antioqueño de Investigación, 2019.
p. 568 -- (Ingeniería y ciencia)

Archivo en formato pdf (15 Mb). -- Incluye referencias bibliográficas al final de cada capítulo.

ISBN 978-958-52333-0-0

1. Ingeniería - Innovaciones tecnológicas - Investigaciones
2. Investigación industrial I. Serna M., Edgar, ed. II. Serie

CDD: 620 ed. 23

CO-BoBN- a1047651

Investigación Científica ISBN: 978-958-52333-0-0
DOI: <http://doi.org/10.5281/zenodo.3387679>

Desarrollo e Innovación en Ingeniería
Serie: Ingeniería y Ciencia
© Editorial Instituto Antioqueño de Investigación

Cuarta Edición: agosto 2019
ISBN: 978-958-52333-0-0
Publicación electrónica gratuita

Copyright © 2019 Instituto Antioqueño de Investigación IAI™. Salvo que se indique lo contrario, el contenido y distribución de esta publicación está autorizado bajo [Creative Commons Licence CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Edición general: Instituto Antioqueño de Investigación IAI
Diseño: IAI, Medellín, Antioquia.
Editorial Instituto Antioqueño de Investigación es Marca Registrada del *Instituto Antioqueño de Investigación*. El resto de marcas registradas pertenecen a sus respectivos propietarios.

La información, hallazgos, puntos de vista y opiniones contenidos en esta publicación son responsabilidad de los autores y no reflejan necesariamente los puntos de vista del Instituto Antioqueño de Investigación IAI, y no se garantiza la exactitud de la información proporcionada en este documento.

Diseño, edición y publicación: Editorial Instituto Antioqueño de Investigación
Instituto Antioqueño de Investigación IAI
<http://fundacioniai.org>
[contacto\(AT\)fundacioniai.org](mailto:contacto(AT)fundacioniai.org)

Presentación

Las diferentes disciplinas ingenieriles avanzan cada año en cuanto a desarrollos y descubrimientos que, con el soporte de la ciencia, le brindan mejores alternativas de vida a la sociedad. Aunque ya sea por mal uso o por intereses comerciales, esas alternativas no siempre logran su objetivo. En todo caso, en esta edición de *Desarrollo e Innovación en Ingeniería* presentamos los resultados del trabajo de una serie de investigadores en áreas como las Ciencias Computacionales e Ingenierías como la Civil, Química, Electrónica, Mecánica, Ambiental e Industrial, y cuyo aporte engrandece al cuerpo de conocimiento de esta disciplina.

El lector podrá encontrar datos, referencias, información y procesos con los que esperamos pueda responder a sus inquietudes en relación con la ingeniería, además de proyectar su trabajo académico e investigativo. Esto es posible gracias a que el contenido de este libro se estructura de forma secuencial por área ingenieril, mostrando los procesos y métodos de investigación que hicieron posible los resultados y que sus autores plasman en cada uno de los capítulos.

Como libro resultado de investigación estamos seguros de que será de gran utilidad para el lector, por la versatilidad del contenido y porque los investigadores y académicos, que publican en esta cuarta edición, se esforzaron al máximo para cumplir con las exigencias de la revisión técnica y, posteriormente, con las observaciones de la revisión científica a la que se sometió el presente texto.

Consiste pues de una obra maravillosa y agradable en su lectura, en la que tanto los profesores como los investigadores y estudiantes encontrarán una fuente de información que, en la medida que analicen y entiendan su contenido, podrán convertir en conocimiento y sacar el mejor provecho del mismo. Los autores de los capítulos estarán atentos a responder sus inquietudes y a entablar, en la medida de lo posible, una relación de trabajo que conlleve a nuevas etapas de trabajo o a nuevas formas de validación de los resultados.

Esperamos entonces que el libro se ajuste a sus necesidades y que se distribuya entre la comunidad, ya que su publicación es libre y, como las ediciones anteriores, se puede descargar sin ninguna restricción desde el sitio de la Editorial Instituto Antioqueño de Investigación.

Contenido

	Pág.
Un modelo de gestión de conocimiento aplicable a las pequeñas y medianas fábricas de software Adriana Maulini, Luciano Straccia, Florencia Pollo C.	5
Revisión de enfoques y comparación de recursos para el análisis de sentimientos en español en Twitter Víctor Rojo, Paola Britos, Florencia Pollo C.	17
Predicción del tráfico de una red inalámbrica basada en redes neuronales artificiales mediante el algoritmo de Levenberg-Marquardt Ramiro Osorio D., Martha Y. Segura R., Mauricio A. Villalba	27
Estudio de perfiles de los tesis de las carreras de grado de Informática de la Universidad de Morón, Argentina Marisa Panizzi, Iris Sattolo, Javier Lafont, Nicolás Armilla	37
Hacking en servicios Web a través de redes WiFi abiertas Francisco C. Valle D., Carlos S. Herrera G., Cristian S. Pedraza C.	49
Análisis bibliométrico al desarrollo de software: Tendencias de investigación Francisco Clemente V., Paulo A. Gaona G., David A. Guerreo P.	58
Estrategia para conectar las necesidades de la industria con los proyectos de aula de los ingenieros de sistemas: Alianza Globant-Universidad de Medellín Samayda Herrera, Liliana González P., Juan Guillermo F., Mario Luna del R.	74
Una aplicación móvil de Realidad Aumentada para la enseñanza de la gestión de almacenes en logística Juan D. Bernal P., Jorge Bacca, Julio M. Daza E.	85
Diseño de un algoritmo para solución de un problema de secuenciación de N tareas en M máquinas mediante la combinación de las reglas de despacho Earliest Due Date y Longest Processing Time en un ambiente de manufactura flexible Germán Cock S., Harold S. Cardona Z., Pedro D. Medina V.	96
Desarrollo de una herramienta computacional para el análisis de deformación debida a una carga perpendicular en estructuras tipo sándwich fabricadas en material compuesto Daniel Agudelo, Jorge E. Gaitán A., Daniela Sacristán M.	105
Caracterización de las propiedades mecánicas a compresión de estructuras tipo sándwich con núcleo celular manufacturadas bajo técnica de deposición fundida Daniel Agudelo, Fabio A. Merchán R., Sebastián Nieto, Sebastián Rodríguez	120
Una revisión técnico-ambiental de la producción de biodiésel a partir de aceite de fritura residual en Colombia Juan C. Acevedo P., Luis C. Becerra O., Astrid Z. Acevedo R., Fausto R. Posso R.	132
Aplicación del método de balance de energía metal/molde para la determinación del calor latente de solidificación y la evolución de fracción sólida de una aleación Al-7.5%Si Carlos González R., Adrián M. Amaro V.	144
La importancia de los catalizadores en la gasificación de biomasa: Una revisión de la literatura Juan C. Acevedo P., Erika Arenas C., Zulamita Zapata B., Fausto R. Posso R.	153
Efectos de la radiación ultravioleta en la guadua Olga L. Vanegas A., Flor N. Díaz P., Luis Francisco P.	172
Aprovechamiento de condensados de aires acondicionados en climas cálidos para riego automatizado de jardines impulsado por energía solar José C. Galindo G., Carlos E. Arrieta G., Iván Mora U., Iván Maestre A.	177
Vigilancia tecnológica para orientar la enseñanza de la geometría en Ingeniería Civil con apoyo de TIC John J. García M., Sonia J. Moreno J.	195
Análisis del comportamiento a largo plazo de un puente viga cajón construido por voladizos sucesivos en Colombia a partir de deflexiones medidas en campo Luis F. Rincón P., Álvaro Viviescas J., Gustavo Chio C., Edison Osorio B., Carlos A. Riveros J.	203
Comportamiento sísmico de muros delgados de concreto reforzado en zona de amenaza sísmica alta Érica L. Arango P., William Valencia M., Álvaro Viviescas J., Oscar J. Begambre C., Gustavo A. Ospina I.	211
Principios metodológicos para el desarrollo de un plan tecnológico de renovación de balasto de una vía ferroviaria: Revisión sistemática de la literatura Jhoan A. Navarro T., Leonel F. Castañeda H., Silvana Montoya N.	219
Localización automática de vehículos de transporte escolar y monitoreo y control de ingreso y salida de estudiantes mediante la integración de tecnologías GPRS, GPS, Arduino y RFID Félix R. Gómez D., Edgar F. Rodríguez V., Guillermo F. Valencia P.	230
Metodologías para la construcción de un modelo prospectivo para la implementación de redes móviles 5G en Colombia Félix R. Gómez D., Guillermo F. Valencia P., Edgar F. Rodríguez V.	240
Sistema IoT para medición y monitoreo de variables medioambientales en la agricultura inteligente Arnold López S., Christian D. Bohórquez M., Ricardo C. Gómez V.	251

Plan de acción para un Laboratorio de Soluciones Tecnológicas a nivel de educación superior: Caso facultad de ingeniería de la Universidad de Medellín Liliana González P., Mario A. Luna del R., Mauricio González P., Jenny Cuatindioy I.	260
Tendencias investigativas en servicios en redes de telecomunicaciones Carlos A. Gómez R., José L. Arciniegas H.	269
Navegación de un robot móvil autónomo aplicado a la agricultura de precisión: Principios y aplicaciones de geolocalización en un cultivo de prueba Ángel A. Rodríguez A., John A. Figueredo L., Juan A. Chica G.	280
Estudio de eficiencia del alumbrado público en la ciudad de Villavicencio Jefferson Parra C., Pedro L. Muñoz O., Luis M. Pulido O.	290
Caracterización y monitoreo de la radiación solar, índice UV y temperatura en el parque metropolitano María Lucía, vereda La Llanerita, Apiay, Villavicencio Pedro L. Muñoz O., Luis M. Pulido O., Jefferson Parra C.	303
Segmentación y análisis de imágenes termográficas para la detección de cáncer de mama Diannys Granadillo B., eeYaileth Morales, Cesar Torres M., Eberto Benjumea	316
Implementación de tecnologías basadas en Inteligencia Artificial para el análisis de la conducta de robots móviles autónomos en espacios de configuración en colisión Alejandro Armando H., Emanuel Maximiliano A.	324
Modelos de madurez de la gestión de la seguridad y salud en el trabajo Yuber L. Rodríguez R., Harold Hernández C., Magda V. Monroy S.	336
Referentes conceptuales de transferencia de conocimiento y tecnología en Instituciones de Educación Superior: Una revisión sistemática de la literatura Neira Y. Figueroa G., Lizeth F. Serrano C., Hugo F. Castro S.	347
Análisis de tendencias investigativas asociadas a la Industria 4.0 Mary A. Mendoza P., Lizeth F. Serrano C., Ivanhoe Rozo R.	358
Entendiendo el concepto de Innovación Social y las tendencias investigativas asociadas: Revisión de literatura Diana K. Parrado H., Lizeth F. Serrano C., Hugo F. Castro S.	370
Nivel de implementación de prácticas Gestión de la Calidad Total en pequeñas empresas colombianas Stanley D. Álvarez E., Daniel A. La Rotta F.	381
Tendencias de la investigación en producción y logística: Un análisis bibliométrico Lindsay Álvarez P., William Camilo R., Alejandro Guerrero P.	392
Tendencias en la toma de decisiones en ingeniería de servicios Eduyn López S., William C. Rodríguez V., Alejandro Guerrero P.	401
Evaluación de una metodología de generación de ideas de negocio basado en patentes libres Jonathan D. Morales M., Eduwin A. Flórez O., Marta E. Carreño G., Andrea C. Patarroyo R.	422
Identificación de las causas de variación en los procesos de esterilización de un hospital privado a través de la metodología Six SIGMA: Fases definir – medir Ivanhoe Rozo R., Ivon A. Vanegas R., Lizeth F. Serrano C., Mary A. Pérez M., Flor N. Díaz P.	434
Aplicación del ciclo de vida Business Process Management al proceso de poscosecha en una empresa de flores Claudia Y. Rodríguez R., Angie P. Mondragón A.	450
Recubrimiento tipo carburo de Tantalio y carburos de Hafnio con capa intermedia de oro para elevar la resistencia a procesos corrosivos Willian Aperador C., Pablo A. Guillermo G., Jorge H. Bautista R.	458
HYNMERS: Diseño de un sistema mecánico pasivo para rehabilitación de miembros inferiores inspirado en terapia acuática Sergio Zapata V., Manuel J. Betancur, Vera Z. Pérez, Juan C. Yepes, Juan C. Franco M.	469
Tendencias investigativas en ingeniería mecatrónica: Un análisis bibliométrico Hernán González A., Daniel F. Peñarete M., Hugo F. Lobatón G.	482
Caracterización mecánica a tensión de impresiones 3D de PLA y ABS modeladas por deposición fundida Daniel Agudelo, David Díaz, Andrés Ibarra, Carlos Cruel, Miguel Montoya, Camilo Hernández	497
Incidencia de las condiciones atmosféricas sobre tecnologías de enfriamiento húmedo y seco para sistemas de refrigeración de centrales de generación de energía Frank Ortiz G., Andrés F. Colorado G., Sergio C. Agudelo F.	508
Análisis termoeconómico del distrito térmico La Alpujarra en Medellín Juan D. Amariles L., Sergio C. Agudelo F., Andrés F. Agudelo S.	519
Marco legal para la conversión eléctrica de vehículos de combustión interna de servicio público en Colombia: Una perspectiva de sostenibilidad Tatiana C. Fragozo G., Sneyder Pastrán S., Leonel Castañeda	533
Efecto de un postratamiento térmico en un acero microaleado con diversos procesos termomecánicos Itzel Reyes, Adair Jiménez N., Bernardo Campillo	546

Revisión de enfoques y comparación de recursos para el análisis de sentimientos en español en Twitter

Víctor Rojo¹

Paola Britos²

Florencia Pollo C.³

Universidad Tecnológica Nacional – Argentina

Twitter ha surgido como una plataforma en la que a diario se expresan millones de opiniones las cuales representan un enorme valor para empresas y organizaciones. Para explotar este inmenso corpus de opiniones, el análisis de sentimientos se ha planteado como un método para clasificar la polaridad de los mensajes subjetivos que se comparten en la red social. En vista de que los resultados producidos por los clasificadores de sentimientos a menudo dependen, en parte, de la calidad de los recursos utilizados para la tarea de análisis de sentimientos, este trabajo se enfoca en la corpora, léxicos y herramientas de análisis utilizados en la bibliografía; específicamente, los participantes de TASS 2017. Los resultados de este trabajo aportan una revisión de enfoques, así como un análisis comparativo de los recursos utilizados en el estado del arte y apuntan hacia posibles líneas de investigación relacionadas al análisis de sentimientos en Twitter en español.

1. INTRODUCCIÓN

El servicio de *microblogging* Twitter se mantiene como una de las redes sociales más populares para el intercambio de opiniones en torno a una gran variedad de temas. En la actualidad, se contabilizan cerca de 335 millones de usuarios activos al mes, de los cuales 68 millones pertenecen a los Estados Unidos y los 267 millones restantes a la comunidad internacional [1]. A diferencia de métodos más tradicionales, recolectar y analizar datos de la plataforma es una alternativa que permite encuestar un amplio número de participantes con un menor número de recursos [2]. Por esta razón, áreas relacionadas a las ciencias políticas, económicas, sociales y de investigación de mercado estudian Twitter activamente con un interés especial en las estadísticas agregadas que surgen de los millones de mensajes (o tweets) producidos por sus usuarios todos los días [3]. El conjunto de las opiniones extraídas durante la ejecución de las técnicas resulta de gran valor para empresas, gobiernos y demás organizaciones por sus posibles aplicaciones, entre las cuales se encuentran el delinear perfiles, conocer sentimientos e ideas de futuros consumidores o votantes, relevar expectativas y realizar predicciones. A menudo, el fin de estos procesos es auxiliar en la toma de decisiones estratégicas que se alineen con los objetivos de la organización.

Una de las formas utilizadas para analizar el contenido de los datos de la red social es por medio del análisis de sentimientos, aunque los términos empleados en los trabajos que estudian y exploran el tratamiento computacional de las opiniones, sentimientos y subjetividad en textos incluyen expresiones como minería de opiniones (*opinion mining*), análisis de subjetividad (*subjectivity analysis*) y minería de reseñas (*review mining*), entre otras [4]. En este sentido, los enfoques de las soluciones actuales al problema del análisis de la información subjetiva abarcan una variedad de técnicas pertenecientes a distintas ciencias, campos y subcampos interdisciplinarios como la Lingüística Computacional, el Procesamiento del Lenguaje Natural (PLN), la Inteligencia Artificial (IA) y la ciencia de los datos, o *Data Science*.

El objetivo principal del análisis de sentimientos consiste en la clasificación automática de textos subjetivos en categorías de polaridad previamente determinadas. Normalmente, el contenido puede ser etiquetado como positivo, negativo o neutral [5], aunque estas agrupaciones pueden variar dependiendo de la granularidad que busque el estudio. En el pasado, el análisis de sentimientos se ha aplicado a distintos medios como blogs, foros y otros documentos, y los retos relacionados a la tarea han sido ampliamente documentados y estudiados. Como indica [6], entre los desafíos comúnmente tratados en el dominio están la separación de cadenas de texto en segmentos para su procesamiento (*tokenization*), eliminación de palabras que no contengan una carga de polaridad (*stop words*), la detección y manejo de negaciones en las oraciones y la presencia de contenido multimodal en estos medios.

Adicionalmente, Giachanou y Crestani [6] señalan que, como resultado de las características propias de la red social, existen elementos particulares que deben ser considerados en el análisis de sentimientos en Twitter. Por ejemplo, aunque fue aumentado recientemente [7], el límite de 280 caracteres impuesto a los mensajes por la plataforma tiene como efecto textos cortos que con regularidad hacen uso de abreviaciones. Lo que complica en mayor medida la tarea es el uso del lenguaje informal por parte de los usuarios para expresar sus opiniones. Debido a esta característica, es posible que en el contenido figuren faltas ortográficas, elongación de palabras para indicar énfasis, uso de *slang*, lenguaje figurado, ironía o sarcasmo y la codificación de emociones en *emoticons*. Por último, a causa de la audiencia internacional con la que cuenta el servicio de microblogging, con frecuencia se pueden encontrar mensajes escritos

¹ vmrojo@gmail.com

² pbritos@unrn.edu.ar

³ fpollo@posgrado.frba.utn.edu.ar

en múltiples idiomas o en una mezcla de los mismos. En conjunto, estas dificultades hacen que el análisis de sentimientos en Twitter sea una tarea no trivial [6]. El formato con el que se presentan los contenidos ante el usuario se puede consultar en la Figura 1, la cual ilustra el primer mensaje que se publicó en el sitio de Twitter y que presenta ejemplos del cuerpo del tweet, contador de réplicas, retweets y favoritos, nombre de usuario y opción para convertirse en seguidor del autor, así como otra información relevante [8].



Figura 1. Ejemplo de un tweet

De la misma forma, en la Figura 2 se puede observar una de las réplicas dirigidas hacia la publicación anterior unos años después. En el nuevo mensaje se pueden apreciar ejemplos de menciones tanto para el autor original (@jack), como para la cuenta oficial del servicio (@twitter), al igual que el uso de hashtags para asociar el comentario a dos temas relacionados a Twitter.



Figura 2. Ejemplo de respuesta a un tweet

Además de la interfaz de usuario, estas características y otra metadata que puede resultar de interés puede ser recuperada por medio del API oficial, después de realizar unos simples pasos de configuración.

2. ENFOQUES DEL ANÁLISIS DE SENTIMIENTOS

La mayoría de las técnicas a menudo empleadas para el reconocimiento de la polaridad de los textos subjetivos pueden ser separadas, *grosso modo*, en tres enfoques: basadas en el léxico, en aprendizaje automático y el abordaje híbrido [8]. Sin embargo, las clasificaciones nombradas no son exhaustivas y en la literatura es posible encontrar soluciones que no se adaptan por completo a uno de los tres paradigmas mencionados. Las características de cada uno de estos enfoques se plantean a continuación.

2.1 Enfoque basado en léxico

Uno de los primeros enfoques que se popularizó entre la comunidad de investigadores dedicada al análisis de sentimientos fue el uso de técnicas basadas en léxico. Los métodos encontrados en esta categoría tratan de realizar las clasificaciones de polaridad de los textos por medio de colecciones de elementos que han sido previamente analizados, a menudo en un proceso manual realizado por humanos con conocimiento de los términos. Es común que en los léxicos se marquen estos elementos con etiquetas que describan la connotación a la que tienden (negativo o positivo, en el caso más básico), o con valores numéricos en una escala predeterminada para indicar tanto su polaridad como su intensidad (-5 a +5, por ejemplo). Existen algunos diccionarios que además realizan una asociación de las palabras a distintos sentimientos, entre los que se pueden encontrar el enojo, la tristeza, el miedo o la sorpresa.

Formalmente, los términos que se recolectan en un léxico se pueden entender como secuencias de uno o varios elementos; secuencias que se componen únicamente por un elemento reciben el nombre especial de unigramas, mientras que a las secuencias que tienen dos y tres elementos se les denomina bigramas y trigramas, respectivamente. Para generalizar, n-grama se refiere entonces a una cadena con n elementos. Si bien un elemento sigue siendo una abstracción, para el caso del procesamiento del lenguaje natural en textos esto puede ser entendido como una letra, sílaba o palabra. Posiblemente por ser una de las opciones más naturales, muchos de los léxicos que se usan para el análisis de sentimientos se pueden definir como diccionarios de palabras, aunque estos no siempre se limitan al uso de unigramas en sus listas.

Habiendo descrito el formato del léxico, realizar el análisis básico de un texto para identificar la polaridad de las opiniones expresadas resulta una tarea sin mucha complejidad. El objetivo en este proceso es comparar el contenido del mensaje con los términos en el diccionario. En este tipo de enfoque, la polaridad se determina por medio de una fórmula que considere los valores listados en el léxico para predecir la carga sentimental del mensaje. Si bien el

enfoque resulta sencillo y directo en comparación a otras técnicas que se describirán en los siguientes puntos, cuenta con algunas desventajas dada la complejidad inherente de los lenguajes y su uso durante la comunicación, así como la naturaleza del ambiente de las redes sociales. Uno de los principales problemas que presenta es su limitada capacidad para manejar negaciones en los mensajes. Por ejemplo, en un mensaje tan simple como “No me gusta”, se puede encontrar una palabra con carga negativa (“No”), así como una positiva (“gusta”). Dependiendo de la evaluación de los valores asignados a cada palabra en el léxico, el resultado del análisis puede variar, sin embargo, para una persona, es claro que el mensaje tiene una connotación negativa.

Algunos casos con uso de negaciones se pueden llegar a tratar haciendo uso de n-gramas mayores a 1 (“no me gusta”), mas no resulta práctico a la hora de considerar todos los casos reales y sus variaciones. Considerando que la utilidad de un diccionario está sujeta a los términos en el mismo, un inconveniente que surge a raíz de esto es la constante necesidad de ser actualizados; especialmente por la informalidad del lenguaje y el uso intensivo de slang que se maneja en las conversaciones entre los usuarios. Por último, los léxicos disponibles para el análisis de sentimientos en Twitter suelen centrarse alrededor de un solo lenguaje, por lo que sería necesaria la incorporación de varios léxicos al analizar mensajes que contengan una mezcla de idiomas.

2.2 Enfoque basado en aprendizaje automático

Otra de las propuestas que los investigadores han desarrollado para el análisis de los textos y la identificación de sentimientos en su contenido se centra alrededor de técnicas relacionadas a la inteligencia artificial utilizadas en conjunto con grandes volúmenes de datos. En este sentido, la tarea resulta particularmente adecuada para este tipo de soluciones debido al rol que redes sociales como Twitter pueden tener como grandes repositorios de datos. Tradicionalmente, el objetivo es planteado en este campo como un problema de clasificación [9]. La aplicación de técnicas estadísticas sobre un set de datos es lo que le permite a una computadora aprender y realizar las predicciones o clasificaciones correspondientes, aunque puede que estas no siempre sean acertadas. La generalización necesaria para la clasificación de los mensajes se da a raíz de una serie ejemplos proporcionados en forma de un *dataset* o corpus en una etapa inicial. A esta etapa, en la jerga del dominio, se le denomina fase de entrenamiento y su correcto desempeño tiene un impacto sobre los resultados del clasificador. Los datos en los conjuntos de entrenamiento tienen características especiales que dependen del tipo de aprendizaje automático empleado. En la bibliografía del análisis de sentimientos en Twitter se pueden encontrar aplicaciones de métodos supervisados, no supervisados y semi-supervisados.

- *Métodos supervisados.* Para el caso del aprendizaje automático con métodos supervisados, el entrenamiento de los modelos se lleva a cabo usando datos que han sido previamente etiquetados. Las etiquetas de polaridad, al igual que en el enfoque basado en léxicos, suelen indicar si el mensaje tiene una connotación positiva o negativa, y pueden extenderse para describir neutralidad, falta de emoción o diversas intensidades. Los procesos de clasificación de los datos de entrenamiento tienden a ser realizados por equipos de humanos que leen, interpretan y asignan la etiqueta idónea para cada mensaje. En vista de que hay distintas regulaciones que se deben contemplar e implementar para lograr un correcto etiquetamiento de los datos, y sumado al hecho de que un set de entrenamiento se puede componer de miles de mensajes, o incluso millones, la confección de estos datasets resulta muy costosa. No obstante, se han explorado soluciones que buscan aminorar el esfuerzo requerido para este tipo de tareas, tales como el uso de técnicas colaborativas o de *crowdsourcing* y los métodos semi-supervisados de aprendizaje automático. En la Figura 3, se muestra una representación gráfica donde se ilustran las distintas etapas que comúnmente componen el proceso de análisis de sentimientos que hace uso del método supervisado de aprendizaje automático según [6].

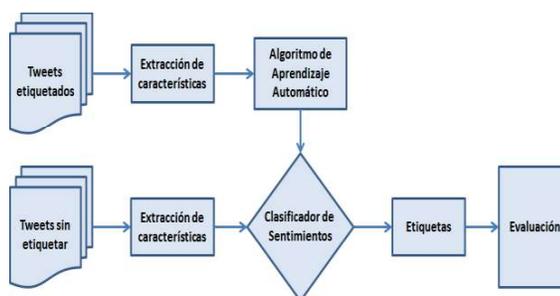


Figura 3. Proceso típico de análisis de sentimientos basado en método supervisado

- *Método semi-supervisado.* Como mencionan Davidov, Tsur y Rappoport [10], el objetivo de este enfoque es tomar ventaja de la vasta cantidad de mensajes, originalmente sin anotaciones de polaridad en su estado natural, para mejorar la eficacia de los clasificadores; esto se logra por medio de un proceso iterativo en el que se combina el uso de datos etiquetados y no etiquetados. Los pasos básicos para llevar a cabo el método semi-supervisado involucran, en una primera instancia, preparar un clasificador con el corpus de mensajes de entrenamiento que incluye etiquetas de polaridad. Una vez finalizado el primer entrenamiento del modelo, se aplica el clasificador

sobre un segundo corpus compuesto por datos no etiquetados para generar probabilidades de clases de sentimiento para cada mensaje. De los resultados se seleccionan las publicaciones con una alta probabilidad para integrarse al corpus de entrenamiento, mientras que el resto se queda en el set no etiquetado. El proceso se puede repetir por un número definido de iteraciones o hasta que el estado del corpus de entrenamiento sea satisfactorio.

- *Métodos no supervisados.* Al igual que los otros enfoques dentro de la familia de los métodos basados en el aprendizaje automático, también hacen uso de textos subjetivos para el entrenamiento de sus clasificadores; a diferencia de los anteriores, en los que se lleva a cabo una selección de características en base a los contenidos de la publicación, en los métodos no supervisados se puede hacer uso de datos sin etiquetar para ganar información contextual de grandes colecciones de mensajes [11]. Mientras que los datos usados para los métodos supervisados suelen ser más escasos por los costos en los que se incurren durante su elaboración, en el enfoque no supervisado se alcanzan nuevas órdenes de magnitud al usar decenas de millones de tweets de fácil acceso y que no requieren de un procesamiento significativo. Este gran volumen de datos es esencial para la aplicación de técnicas como *embeddings* de palabras o caracteres o para proporcionar buenos parámetros de inicialización para los modelos a entrenar, algunos factores que determinan la eficacia del clasificador [12].

Una vez que se ha determinado el set de entrenamiento, cada mensaje etiquetado es procesado para realizar una extracción de características. Se debe prestar atención a su selección, ya que éstas, al igual que el algoritmo de aprendizaje automático elegido para la tarea, tendrán un impacto sobre el desempeño del clasificador resultante.

2.3 Enfoque híbrido

Las soluciones de este tipo y como lo insinúa su nombre, tratan de mejorar el desempeño de los clasificadores automáticos complementando los enfoques descritos hasta el momento por medio de la combinación de dos o más de sus técnicas. Por ejemplo, en este grupo se ubican los trabajos que experimentan con el uso de un clasificador SVM o Naive Bayes, los dos del tipo de aprendizaje automático supervisado, alimentado por puntuaciones realizadas por un proceso en base a léxico [13]. De la misma forma califican aquellos estudios que realizan minería de opiniones por medio de una combinación de SVM y redes neuronales [14], ambos del área del aprendizaje automático supervisado y no supervisado, respectivamente. En los trabajos anteriores, los investigadores reportan resultados superadores en el caso de los métodos híbridos en comparación a las soluciones que se centraron únicamente en resolver el problema con un enfoque, ya sean basadas en léxico o en aprendizaje automático.

2.4 Otros enfoques

En su trabajo [6], Giachanou y Crestani realizan un análisis de varias soluciones que no logran ser encuadradas por completo por uno de los tres enfoques descritos en los puntos anteriores. En esta categoría identifica trabajos novedosos que afrontan el problema con ideas influenciadas por teorías pertenecientes a disciplinas como la psicología y la sociología o por métodos como el análisis formal de conceptos.

3. RESULTADOS

En el análisis de sentimientos en la red social Twitter existen varios recursos que se utilizan en los enfoques descritos para auxiliar en la tarea de clasificación de mensajes. Estos recursos son empleados en las soluciones para el diseño, implementación, entrenamiento y validación de características y clasificadores, por lo que tienen un impacto en los resultados obtenidos. Los recursos en los siguientes puntos fueron identificados tras una revisión de los trabajos presentados en la edición 2017 (**¡Error! No se encuentra el origen de la referencia.**) de la competencia afiliada a la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN): el Taller de Análisis Semántico en la SEPLN (TASS).

Tabla 1. Participantes en TASS 2017 por orden alfabético

Grupo	Trabajo	Ref.
C100TPUCP	C100TPUCP at TASS 2017: Word Embedding Experiments for Aspect-Based Sentiment Analysis in Spanish Tweets	[15]
ELiRF	ELiRF-UPV en TASS 2017: Análisis de Sentimientos en Twitter basado en Aprendizaje Profundo	[16]
GSI	Applying Recurrent Neural Networks to Sentiment Analysis of Spanish Tweets	[17]
INGEOTEC	A Genetic Programming Approach to Sentiment Analysis for Twitter: TASS'17	[18]
ITAINNOVA	FastText como alternativa a la utilización de Deep Learning en corpus pequeños	[19]
JACERONG	Classifier Ensembles that Push the State-of-the-Art in Sentiment Analysis of Spanish Tweets	[20]
LexFAR	LexFAR en la competencia TASS 2017: Análisis de sentimientos en Twitter basado en lexicones	[21]
OEG	OEG at TASS 2017: Spanish Sentiment Analysis of tweets at document level	[22]
RETUYT	RETUYT en TASS 2017: Análisis de Sentimientos de Tweets en Español utilizando SVM y CNN	[14]
SINAI	SINAI en TASS 2017: clasificación de la polaridad de tweets	[23]
Tecnolengua	Tecnolengua Lingmotif at TASS 2017: Spanish Twitter Dataset Classification Combining Wide-coverage Lexical Resources and Text Features	[24]

En la bibliografía se reconocen distintos tipos de recursos para el análisis de sentimientos en Twitter y para este trabajo se ha decidido agruparlos en las siguientes categorías según sus características y función: corpora, léxicos y herramientas de análisis.

3.1 Corpora

Ya sea que se requieran para el desarrollo de los mismos sistemas o durante su evaluación, en los métodos actuales de análisis de sentimientos es indispensable el uso de sets de datos, o corpora. La formación de estos datasets se facilita gracias herramientas en forma de API que proporciona Twitter, con las cuales se pueden recolectar mensajes en base a búsquedas específicas. Los conjuntos de datos luego son anotados, si el enfoque de análisis de sentimientos lo demanda, por medio de una variedad de técnicas que van desde un trabajo manual a distintos grados de automatización. Las dimensiones analizadas para el caso de este tipo de recurso son:

- *Corpus o Dataset.* El nombre por el cual es conocido o referenciado el set de datos. En caso de múltiples nombres, se considera aquel que sea más prominente en la bibliografía analizada.
- *Autores.* El grupo o sujeto responsable por el desarrollo del recurso léxico. Se indica para poder diferenciar aquellos recursos que no fueron proporcionados por los organizadores de la competencia.
- *Método.* Habla de las técnicas utilizadas para etiquetar los tweets que conforman el corpus o, en su defecto, indica la falta de etiquetas en el mismo.
- *Etiquetas.* Enlista las categorías de polaridad en las que los mensajes fueron clasificados o, en su defecto, señala la falta de las mismas.
- *Positivo.* El número de mensajes o entidades que fueron clasificados por su contenido con carga positiva.
- *Neutral.* El número de mensajes o entidades que fueron clasificados por su contenido con carga neutral.
- *Negativo.* El número de mensajes o entidades que fueron clasificados por su contenido con carga negativa.
- *None.* El número de mensajes o entidades que fueron clasificados por la falta de una carga emocional en su contenido.
- *Tamaño.* La cantidad total de elementos que conforman el cuerpo del corpus. Cabe aclarar que este valor no necesariamente refleja la suma de las cantidades que aparecen en las dimensiones relacionadas a la polaridad; existen casos, en corpora de TASS, por ejemplo, en los que únicamente los subsets de entrenamiento han sido etiquetados, por lo que el tamaño total del corpus es mayor a la cantidad de mensajes anotados.

Los *datasets* encontrados en la bibliografía de los trabajos en TASS 2017, por orden alfabético, son se presentan en la Tabla 2.

Tabla 2. Comparación de *datasets* en español identificados en TASS 2017 por orden alfabético

Corpus/Dataset	Autores	Método	Etiquetas	Positivo	Neutral	Negativo	None	Tamaño
DS dataset	Moctezuma et al.	Distant Supervision	P, N	9 millones	-	9 millones	-	18 millones
General Corpus / General TASS	TASS	Semi automático	P+, P, NEU, N, N+, NONE	2,884	670	2,182	1,483	68,017
General Corpus 1K	TASS	-	Sin etiquetas	-	-	-	-	1,000
InterTASS	TASS	Manual	P, NEU, N, NONE	1,116	418	1,404	475	3,413
Mozetič, Grčar & Smailovič corpus	Mozetič, Grčar, Smailovič	Manual	P, NEU, N	69,571	54,017	16,472	-	140,060
RETUYT corpus	RETUYT	Unión	P, NEU, N, NONE	3,358	872	2,819	1,684	8,733
Social_TV	TASS	Manual	P, NEU, N	1,547	999	1016	-	2,773
Spanish Billion Words Corpus and Embeddings	Cardellino	-	Sin etiquetas	-	-	-	-	1,500 millones
STOMPOL	TASS	Manual	P, NEU, N	218	239	586	-	1,284

3.2 Léxicos de Sentimientos

Para los léxicos lingüísticos se han tomado como punto de partida las dimensiones utilizadas por Hinojosa et al. [25]. Además, se ha considerado pertinente agregar información sobre el proceso de creación de los léxicos estudiados ya que estos podrían tener un impacto en el desempeño del recurso; por ejemplo, se ha señalado que en los casos en los que un proceso de traducción automática ha sido usado en un léxico, el resultado de los clasificadores puede depender de la calidad de la traducción [26]. Específicamente, se agregan dimensiones para indicar las fuentes que fueron utilizadas para formar el contenido lingüístico del léxico, los métodos que se aplicaron para su creación y herramientas o agentes que hayan auxiliado durante el proceso. Las dimensiones seleccionadas para la comparación de los recursos léxicos se definen a continuación:

- *Léxico.* El nombre por el cual es conocido el recurso. En el caso de recursos que se conocen por varios nombres, se trataron de incluir aquellos que fueron usados tanto en el trabajo estudiado en TASS como en la publicación original del recurso léxico.
- *Idioma.* El idioma o idiomas en los que está disponible el léxico. Debido a que los recursos fueron utilizados para el análisis de sentimientos en español, las demás dimensiones se enfocan únicamente en las características de esta versión; de no tener una versión en español, se considera el lenguaje al que se le haya aplicado la traducción, según el trabajo que lo usó.
- *Método y Ejecución.* El tipo de método que fue utilizado para la elaboración del léxico y la forma específica en la que fue ejecutado el método de elaboración. Puede referirse una herramienta o técnica, entre otras cosas.
- *Base.* Las fuentes que fueron consultadas o integradas para la confección de las listas de términos en el léxico.
- *Dominio.* El dominio de los términos en el recurso lingüístico. Los valores se pueden ver influenciados dependiendo del contexto.
- *Tamaño.* La cantidad de términos encontrados en el recurso léxico
- *n-grams.* El número de elementos encontrados en las secuencias de los términos.
- *Categorías.* Los nombres de las etiquetas que fueron usadas para categorizar los términos del recurso léxico.
- *Escala.* La escala utilizada para denotar intensidad de la emoción anotada para cada término en el diccionario.

En los trabajos estudiados se identificaron 11 recursos léxicos usados en las soluciones de los equipos participantes. Los resultados de esa revisión fueron recopilados y ordenados en la Tabla 3.

Tabla 3. Comparación de recursos léxicos identificados en TASS 2017 por orden alfabético

Léxico	Idioma	Método y Ejecución	Base	Dominio	Tamaño (term.)	n-grams	Categorías	Escala
AFINN-111	Inglés	Manual (autor)	Palabras obscenas Tweets de COP15 Original Balanced Affective Word List Urban Dictionary The Compass DeRose Guide to Emotion Words	General Slang United Nation Climate Conference (COP15)	2,477	1, 2, 3	Positivo Negativo	-5 a +5
ANEW	Inglés	Manual (alumnos)	150 palabras de Mehrabian & Russell (1974) 450 palabras de Bellezza, Greenwald, & Banaji (1986)	General	1,030	1	Placer Excitación Dominación	9 puntos
Dictionary of Affect in Language	Inglés	Manual (voluntarios)	Diccionario Listas de Conte, Plutchik y Russell Selección de un diccionario en Inglés	Palabras emocionales	8,742	1	Agrado Excitación Evocación de imágenes	3 puntos
ElhPolar	Español	Traducción Semi-automática (Diccionario Inglés-Español + Refinamiento)	Opinion Finder lexicon Training corpus (TASS)	General Turismo	4,511 (1,654 pos, 2,857 neg)	1, 2, 3, 4, 5	Positivo Negativo	Binario
iSOL	Español	Traducción Automática (Reverso)	BLEL (Bing Liu)	General Cinema (eSOL)	8,133 (2,509 pos, 5,626 neg)	1	Positivo Negativo	Binario
LexFAR	Español		InterTASS corpus	General	9,058	1		
MADS / Hinojosa corpus	Español	Manual (Survey Monkey)	LEXESP Trabajos pasados de los autores	General	875	1	Valencia Excitación Concreción Felicidad Enojo Tristeza Miedo Aversión	9 puntos 5 puntos
ML-SentiCon	Inglés Español Catalán Vasco Gallego	Traducción Automática	SentiWord Net 3.0	General	11,302	1, 2, 3, 4, 5, 6	Positivo Negativo	-1 a +1

NRC / EmoLex	Inglés Otros idiomas	Inglés Manual (Amazon's Mechanical Turk) Otros Idiomas Traducción Automática (Google Translate)	Macquarie Thesaurus General Inquirer WordNet Affect Lexicon	General	14,182 palabras ~25,000 senses*	1	Sentimientos Positivo Negativo Emociones Ira Anticipación Alegría Confianza Miedo Sorpresa Tristeza Aversión	Binario
RETUYT	Español	Combinación (Expansión de lemas con FreeLing)	ML-SentiCon ElhPolar Spanish SO-CAL		4,730		Positivo Negativo	
Strength Lex	Español	Transferencia (Framework)	Opinion Finder lexicon Spanish WordNet WordNet SentiWordNet	General	1,347 (477 pos, 870 neg)	1	Positivo Negativo	Binario

3.3 Herramientas de Análisis

Por último, se ha compilado también una herramienta comparativa para los recursos en la categoría de herramientas de análisis. En esta instancia se han considerado dimensiones que expresen las características del tipo de funciones que cada recurso realiza, intentando que estas se compartan en los distintos tipos de herramientas encontradas. Las dimensiones resultantes para la comparación de los recursos de análisis son las siguientes:

- *Herramienta.* El nombre que recibe la herramienta o por el cual es conocida
- *Tecnología.* Indica de qué forma o para qué lenguajes de programación está disponible la herramienta para los usuarios. Normalmente, coincide con el lenguaje de la implementación
- *Categoría.* Identifica la categoría a la que pertenece la herramienta en relación a las funciones de análisis que desempeña
- *Análisis.* Especifica los tipos de análisis para los que la herramienta está preparada. Esta dimensión se relaciona a la categoría y los elementos listados en estos campos deben pertenecer a la dimensión señalada por la dimensión anterior
- *Modelo.* Aplica únicamente para cierto tipo de categorías. Detalla el modelo que sigue la implementación del análisis
- *Idioma.* La lista de idiomas que soporta la herramienta para desempeñar los análisis que brinda.

En los trabajos se han identificado 16 herramientas que han sido utilizadas por los participantes del taller TASS para realizar las tareas de análisis de sentimientos en español. Los recursos y sus características se presentan a continuación en la Tabla 4.

Tabla 4. Comparación de herramientas de análisis identificadas en TASS 2017 por orden alfabético

Herramienta	Tecnología	Categoría	Análisis	Modelo	Idioma
B4MSA	Librería (Python)	Implementación de Algoritmos	Clasificador	SVM	Multilingüe
EvoDAG	Librería (Python)	Implementación de Algoritmos	Clasificador	Algoritmo Genético	-
FastText	Librería (C++)	Representación de Palabras Herramienta de Análisis de Sentimiento	Vectores de Palabras Análisis de Sentimientos	NN (para vectores)	Multilingüe
FreeLing	Librería (C++)	Paquete PLN	Análisis Morfológico Reconocimiento de Entidades Nombradas POS tagger Desambiguación Lingüística Etiquetado de Roles Semánticos	-	inglés, español, portugués, italiano, francés, alemán, ruso, catalán, gallego, croata, esloveno, otros
IBM Watson Natural Language Understanding	Cloud API	Paquete PLN	Identificación de Categorías, Conceptos, Emociones, Entidades, Palabras clave, Metadatos, Relaciones, Roles Semánticos,	-	inglés (default), árabe, chino (simplificado), holandés, inglés, francés, alemán, italiano, japonés, coreano, portugués, ruso, español, sueco (depende de función)

Sentimientos					
IXA-pipes	Librería (Java)	Implementación de Algoritmos	POS tagger, Lematizador, Constituent Extractor,	-	vasco, holandés, inglés, francés, gallego, alemán, italiano y español (depende de función)
Keras	Librería (Python)	Implementación de Algoritmos	Red Neuronal	CNN, RNN	-
Lingmotif Learn	Aplicación de Escritorio	Herramienta de Análisis de Sentimiento	Análisis de Sentimientos Visualizaciones	-	inglés, español
Moriarty®	Web API	Transformaciones de Textos	Lematizador Stop words	-	inglés, español
NLTK	Librería (Python)	Paquete PLN	Clasificadores Tokenizador Stemming, POS tagger Otros	-	Inglés
scikit-learn	Librería (Python)	Implementación de Algoritmos	Clasificadores Regresiones Clustering Reducción de Dimensionalidad	Varios	-
TensorFlow	Librería (Python, C)	Implementación de Algoritmos	Red Neuronal	NN	-
TreeTagger	Scripts	Paquete PLN	POS tagger Lematizador	-	Soporte para español
Tweetmotif	Librería (Python, Java)	Motor de Búsquedas	Búsquedas Identificación de Temas Resumen Tokenizador	-	-
Weka	Aplicación de Escritorio	Implementación de Algoritmos	Implementaciones de Aprendizaje Automático Funciones para Minería de Datos Visualizaciones	Varios	-
Word2Vec	Depende (en TASS: Python, C)	Representación de Palabras	Vectores de Palabras	Word2Vec	-

4. CONCLUSIONES

En el presente trabajo se ha llevado a cabo una revisión del estado del arte en el dominio del análisis de sentimientos en español, específicamente en el ambiente de la red social y servicio de microblogging de Twitter, en base a los trabajos presentados en la edición 2017 del Taller de Análisis Semántico en la SEPLN (TASS).

La exploración del análisis de sentimientos en Twitter en lenguajes diferentes al inglés es una necesidad que ha sido reconocida por investigadores dentro de la comunidad hispanohablante, así como aquellos que no forman parte de ella [27]. El desafío de la clasificación de tweets se trata de afrontar por medio de recursos que ayuden en las tareas de transformación de texto y extracción de características. Debido a que la mayoría de los trabajos en el dominio se enfocan en solucionar el problema considerando únicamente un idioma, los recursos creados varían en calidad y cantidad al pasar de un lenguaje a otro. Al contar con el mayor número de herramientas en inglés, la solución de algunos investigadores ha sido, en ocasiones, optar por traducciones automáticas de recursos léxicos para adaptarlos al análisis de tweets en español [8], [26]; sin embargo, como han señalado algunos autores, los resultados pueden variar dependiendo de la calidad de la traducción automática [26]. Es por esta razón que los investigadores con foco en el análisis de sentimientos en Twitter en español han puesto en evidencia en varios estudios [28], [29] la falta de recursos dedicados a este lenguaje.

En las categorías de recursos discutidas en este trabajo se pueden apreciar puntos de mejora en lo relacionado a los recursos para el análisis de sentimientos en español en Twitter. En el área de los sets de datos de entrenamiento, al considerar el dominio de los mismos, se observa la falta de corpora compuesta en su mayoría o totalidad por mensajes con un contenido sarcástico o con uso de negaciones. Aunque a menudo se mencionan entre una de las complicaciones más comunes a la hora de realizar ejercicios de análisis de sentimientos, los datasets estudiados no resaltan la incorporación de mensajes con estas características durante su elaboración. Esta misma área de vacancia ha sido señalada por otros investigadores anteriormente [29].

Relacionado a los recursos léxicos utilizados ampliamente en las propuestas de los investigadores, se observa que pocos recursos de este tipo toman en cuenta el manejo del slang entre los elementos que los conforman. De los recursos estudiados, únicamente el léxico ANEW y el léxico ElhPolar expresan haber usado una fuente para agregar términos de slang de Internet. De la misma forma, en el trabajo de Cruz et al. [30] se comenta que el léxico de Bing Liu (BLEL) agrega términos informales a la composición del léxico. Si bien estos recursos son reconocidos y utilizados

para el análisis de sentimientos, dos de ellos fueron diseñados originalmente para el idioma inglés; por esta razón, se justifica un trabajo similar que contemple el uso de slang para el español.

Tomando en cuenta que los resultados de los clasificadores pueden variar ampliamente dependiendo de la calidad de los recursos léxicos y las herramientas para el pre-procesamiento de los datos involucrados, se considera pertinente focalizar la atención y el esfuerzo en el área de los recursos para el análisis de sentimientos. La falta de variedad en los recursos es una problemática que ha sido identificada por parte de la comunidad dedicada al estudio del análisis de sentimientos en Twitter, por lo que la creación de nuevos recursos exclusivos para el español podría reducir la brecha entre los materiales disponibles para el español y lenguajes como el inglés, los cuales cuentan con un gran número de diccionarios, corpora y otras herramientas de calidad para el desarrollo de la tarea.

REFERENCIAS

- [1] Twitter Investor Relations, «Twitter,» 27 Julio 2018. [En línea]. Available: <https://twitter.com/i/moments/1022804623717875712>. [03 Octubre 2018].
- [2] L. Karami, S. Bennett and X. He, "Mining Public Opinion about Economic Issues: Twitter and the U.S. Presidential Election," *International Journal of Strategic Decision Sciences*, vol. 9, no. 1, pp. 18-28, 2018.
- [3] S. Rosenthal, N. Farra and P. Nakov, "SemEval-2017 Task 4: Sentiment Analysis in Twitter," in *11th International Workshop on Semantic Evaluation (SemEval-2017)*, 2017.
- [4] N. Pang y L. Lee, «Opinion Mining and Sentiment Analysis,» *Foundations and Trends® in Information Retrieval*, vol. 2, n° 1-2, pp. 1-135, 2008.
- [5] M. A. Escortell Pérez, M. Giménez Fayos and P. Rosso, "El Impacto de las Emociones en el Analisis de la Polaridad en Textos con Lenguaje Figurado en Twitter," *Procesamiento del Lenguaje Natural*, no. 58, pp. 85-92, Marzo 2017.
- [6] A. Giachanou and F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," *ACM Computing Surveys (CSUR)*, vol. 49, no. 2, Noviembre 2016.
- [7] Rosen and I. Ihara, "Twitter Blog," Septiembre 2017. [Online]. Available: https://blog.twitter.com/official/en_us/topics/product/2017/Giving-you-more-characters-to-express-yourself.html.
- [8] J. Dorsey, «Twitter,» 21 Marzo 2006. [En línea]. Available: <https://twitter.com/jack/status/20>. [07 Agosto 2018].
- [9] J. Wehrmann, W. Becker, H. E. L. Cagnini and R. C. Barros, "A Character-Based Convolutional Neural Network for Language-Agnostic Twitter Sentiment Analysis," in *Neural Networks (IJCNN), 2017 International Joint Conference*, Anchorage, AK, USA, 2017.
- [10] L. F. Hurtado Oliver, F. Pla and D. Buscaldi, "ELiRF-UPV en TASS 2015: Análisis de Sentimientos en Twitter," in *TASS workshop at SEPLN 2015*, 2015.
- [11] Davidov, O. Tsur y A. Rappoport, «Semi-supervised recognition of sarcastic sentences in Twitter and Amazon,» de *CoNLL '10 Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, 2010.
- [12] P. Segura, A. Quirós and P. Martínez, "Exploring Convolutional Neural Networks for Sentiment Analysis of Spanish Tweets," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, Valencia, España, 2017.
- [13] A. Severyn y A. Moschitti, «Twitter Sentiment Analysis with Deep Convolutional Neural Networks,» de *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Santiago, Chile, 2015.
- [14] O. Kolchyna, T. T. P. Souza, P. C. Treleaven and T. Aste, "Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination," *Handbook of Sentiment Analysis in Finance*, 2015.
- [15] A. Rosá, L. Chiruzzo, M. Etcheverry y S. Castro, «RETUYT en TASS 2017: Análisis de Sentimientos de Tweets en Español utilizando SVM y CNN,» *Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN*, pp. 77-83, 2017.
- [16] F. Tume Fiestas y M. A. Sobrevilla Cabezudo, «C100TPUCP at TASS 2017: Word Embedding Experiments for Aspect-Based Sentiment Analysis in Spanish Tweets,» *TASS 2017: Workshop on Semantic Analysis at SEPLN*, pp. 85-90, 2017.
- [17] L. F. Hurtado, F. Pla y J. González, «ELiRF-UPV en TASS 2017: Análisis de Sentimientos en Twitter basado en Aprendizaje Profundo,» de *Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN*, 2017.
- [18] O. Araque, R. Barbado, J. F. Sánchez-Rada y C. A. Iglesias, «Applying Recurrent Neural Networks to Sentiment Analysis of Spanish Tweets,» de *Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN*, 2017.
- [19] Moctezuma, M. Graff, S. Miranda, E. S. Tellez, A. Coronado, C. N. Sánchez and J. Ortiz, "A Genetic Programming Approach to Sentiment Analysis for Twitter: TASS'17," in *Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN*, 2017.
- [20] R. M. Montañés, R. del Hoyo, J. Veá, R. Aznar and F. J. Lacueva, "FastText como alternativa a la utilización de Deep Learning en corpus pequeños," in *Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN*, 2017.
- [21] A. Cerón, «Classifier Ensembles that Push the State-of-the-Art in Sentiment Analysis of Spanish Tweets,» de *Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN*, 2017.
- [22] A. Reyes, F. Paniagua, B. Priego y M. Tovar, «LexFAR en la competencia TASS 2017: Análisis de sentimientos en Twitter basado en lexicones,» *TASS 2017: Workshop on Semantic Analysis at SEPLN*, pp. 51-57, 2017.
- [23] L. Navas y V. Rodríguez, «OEG at TASS 2017: Spanish Sentiment Analysis of tweets at document level,» de *Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN*, 2017.
- [24] C. García, A. Montejo-, M. C. Díaz y S. M. Jiménez, «SINAI en TASS 2017: clasificación de la polaridad de tweets,» de *Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN*, 2017.
- [25] A. Moreno y C. Pérez, «Tecnolengua Lingmotif at TASS 2017: Spanish Twitter Dataset Classification Combining Wide-coverage Lexical Resources and Text Features,» de *Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN*, 2017.
- [26] J. A. Hinojosa, N. Martínez, C. Villalba, U. Fernández, A. Sánchez, M. A. Pozo and P. R. Montoro, "Affective norms of 875 Spanish words for five discrete emotional categories and two emotional dimensions," *Behavior research methods*, vol. 48, no. 1, pp. 272-284, 2016.

- [27] Vilares, M. A. Alonso and C. Gómez, "Supervised Sentiment Analysis in Multilingual Environments," *Information Processing & Management*, vol. 53, no. 3, pp. 595-607, Mayo 2017.
- [28] P. Nakov, "Semantic Sentiment Analysis of Twitter Data," in *Encyclopedia on Social Network Analysis and Mining (ESNAM)*, 2017.
- [29] Sidorov, S. N. Galicia and V. A. Camacho, "Construcción de un corpus marcado con emociones para el análisis de sentimientos en Twitter en español," *Revista Escritos BUAP*, vol. 1, no. 1, 2016.
- [30] S. M. Jiménez, M. T. Martín, E. Martínez and L. A. Ureña, "Studying the Scope of Negation for Spanish Sentiment Analysis on Twitter," *IEEE Transactions on Affective Computing*, 2017.
- [31] L. Cruz, J. A. Troyano, B. Pontes y F. J. Ortega, «Un lexicón multilingüe de polaridades semánticas a nivel de lemas,» *Procesamiento del Lenguaje Natural*, nº 53, 2014.