

# LexFAR en la competencia TASS 2017: Análisis de sentimientos en Twitter basado en lexicones

## *LexFAR at TASS 2017: Lexicon-based sentiment analysis in Twitter*

José A. Reyes-Ortiz<sup>1</sup>, Fabián Paniagua-Reyes<sup>1</sup>, Belém Priego<sup>1</sup>, Mireya Tovar<sup>2</sup>

<sup>1</sup> Universidad Autónoma Metropolitana, Departamento de Sistemas  
Azcapotzalco, México

<sup>2</sup> Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación  
Puebla, México

{jaro, al2112002241,abps}@azc.uam.mx, mtovar@cs.buap.mx

**Resumen:** Este artículo describe el sistema para el análisis de sentimientos en Twitter a nivel de un mensaje sometido a la tarea 1 del TASS 2017: Taller sobre Análisis Semántico en SEPLN, desarrollado por el equipo de investigación de la Universidad Autónoma Metropolitana, unidad Azcapotzalco en colaboración con la Benemérita Universidad Autónoma de Puebla. El sistema propuesto utiliza aprendizaje automático, el algoritmo de máquinas de soporte vectorial y lexicones de polaridades semánticas a nivel de lemas para el español. Las características extraídas de los lexicones son representadas mediante el modelo de bolsa de palabras y son ponderadas utilizando la frecuencia de los términos, la cual expresa la ocurrencia del lema en cada tweet. La experimentación muestra resultados prometedores con el uso de lexicones para el análisis de sentimientos a nivel de tweet.

**Palabras clave:** Análisis de sentimientos, aprendizaje automático, máquinas de soporte vectorial, lexicones.

**Abstract:** This paper describes our system for sentiment analysis in Twitter at Tweet level submitted to task 1 of the TASS-2017: Workshop on Semantic Analysis at SEPLN, developed by team of the Autonomous Metropolitan University, Azcapotzalco in collaboration with the Autonomous University of Puebla. The proposed system uses machine learning, vector support machines algorithm and lexicons of semantic polarities at the level of lemma for Spanish. Features extracted from lexicons are represented by the bag-of-words model and they are weighted using TF “Term Frequency” measure, which expresses the occurrence of lemmas in each tweet. Experimentation shows promising results with the use of lexicons for the sentiment analysis in Twitter.

**Keywords:** sentiment analysis, machine learning, support vector machine, lexicons.

## 1 Introducción

El Taller sobre Análisis Semántico en SEPLN (TASS) en su edición 2017, Martínez-Cámara et al. (2017), ha centrado su interés en dos tareas: el análisis de sentimientos a nivel de tweet y el análisis de sentimientos basados en aspectos. Específicamente, la tarea consiste en determinar la polaridad de un tweet, considerando cuatro etiquetas (P, N, NEU, NONE). La idea es evaluar los sistemas de clasificación de polaridad en cuanto a la

precisión para identificar la intensidad de una opinión expresada en cada tweet. Por otro lado, la tarea 2 consiste en clasificar la polaridad a nivel de aspectos. Los sistemas enviados para esta tarea deben de ser capaces de clasificar una opinión dado un aspecto, en una etiqueta de tres niveles de intensidad: Positiva, Negativa y Neutra.

El Análisis de Sentimientos (AS) es una tarea que se ha desarrollado ampliamente para el inglés. Sin embargo, el AS para el español ha sido poco abordado y por ello se tiene una carencia de recursos de análisis de opiniones

para este idioma. El Taller TASS a lo largo de su existencia ha propuesto dicha tarea para el español, en la cual se han propuesto sistemas con diversos enfoques. El aprendizaje profundo, como en (Vilares et al. 2015), (Díaz-Galiano y Montejo-Ráez, 2015) y (Montejo-Ráez y Díaz-Galiano, 2016), ha sido utilizado para representar los tweets y desempeñar el análisis de sentimientos a nivel de sentencias. El uso de vectores de características es ampliamente utilizado en la tarea de análisis de sentimientos. En (Murillo y Raventós, 2016) usan vectores de características de baja dimensión para representación del texto, ellos proponen un modelo simple fundamentado en la normalización de texto con identificación de marcadores de énfasis, el uso de modelos de lenguaje para representar las características locales y globales del texto, y características como emoticones y partículas de negación; (Martínez-Cámara et al., 2015), del mismo modo, construyen vectores de palabras a partir de la información de opinión de recursos lingüísticos; Quirós, Segura-Bedmar, y Martínez (2016), representan los tweets por medio de vectores de palabras ponderados con TF-IDF y son clasificados utilizando algoritmos como máquinas de soporte vectorial (SVM) y regresión logística.

Las aproximaciones híbridas para el análisis de sentimientos, también han sido empleadas, Álvarez-López et al. (2015) desarrollan una aproximación híbrida para el análisis de sentimiento global en Twitter, mediante el uso de clasificadores y aproximaciones sin supervisión, construidas mediante léxicos de polaridad y estructuras sintácticas.

Este artículo se centra en la tarea 1 del TASS 2017, para la cual presenta el sistema del equipo LexFAR de la Universidad Autónoma Metropolitana, unidad Azcapotzalco en colaboración con la Benemérita Universidad Autónoma de Puebla. El sistema propuesto utiliza un enfoque basado en aprendizaje automático, utilizando el algoritmo de SVM y dos lexicones de polaridades semánticas a nivel de lemas para el español. Los lemas de los lexicones son representados mediante el modelo de bolsa de palabras y ponderadas utilizando la medida numérica de *Frecuencia del Término (FT)*, la cual expresa la ocurrencia del lema en cada tweet.

El resto del artículo está organizado como sigue. Primero, se presentan las diversas aproximaciones del sistema para el análisis de

sentimientos a nivel de tweets. Posteriormente, los resultados de la experimentación son expuestos y analizados. Finalmente, las conclusiones y el trabajo futuro son presentados.

## 2 Descripción del sistema propuesto

El sistema propuesto se enfoca en la tarea 1 del TASS en su edición 2017. Para esta tarea, se proporcionaron diversos corpora: InterTASS corpus presentado en (Martínez-Cámara et al., 2017) y dos corpora generales de TASS 2016 para el entrenamiento, los cuales se describen en (García-Cumbreras et al., 2016).

De esta manera, esta sección presenta los elementos que forman parte del sistema sometido a la competencia TASS 2017. Primero, se describen los lexicones utilizados para la clasificación de la polaridad a nivel de un tweet. Posteriormente, se expone la etapa de pre-procesamiento de los tweets y de los lexicones. Después, se expone la representación de los lemas del lexicon y la métrica de ponderación en cada tweet, con la finalidad de obtener los vectores prototipo con las características para el conjunto de entrenamiento y para el conjunto de pruebas. Finalmente, la etapa de identificación de polaridad llevada a cabo con el algoritmo de clasificación máquinas de soporte vectorial.

### 2.1 Lexicones de polaridad

Para el análisis de sentimientos a nivel de oraciones es indispensable el análisis de las palabras del tweet completo para determinar su polaridad. Por ello, en este sistema utilizamos dos lexicones de polaridad con la finalidad de determinar la carga emocional que imprimen las palabras de cada tweet, es decir, que tan positivo, negativo, neutro o sin polaridad es el contenido del tweet.

El primer lexicon llamado *ML-SentiCon* (Cruz et al., 2014), está constituido por 11302 entradas o lemas en español. Este recurso contiene lemas polarizados, con valores que van desde -1.0 “negativo” hasta +1.0 “positivo” y, adicionalmente, un valor de desviación estándar que refleja la ambigüedad resultante del cómputo de la polaridad a partir de los valores de los distintos significados posibles del lema. Además, se tiene la categoría gramatical para cada lema: verbo (v), sustantivo (s), por mencionar algunos. Este lexicon es generado de manera automática utilizando una versión

mejorada del método usado para construir SentiWordNet 3.0 (Baccianella, Esuli, y Sebastiani, 2010). Las evaluaciones de este lexicón han demostrado resultados prometedores.

El segundo lexicón llamado *iSOL* (Lexicón Mejorado de Opiniones en Español), creado por Molina-González et al. (2013), es una versión mejorada del recurso original llamado SOL, el cual proviene de una traducción del lexicón *BLEL*. Diversas mejoras fueron llevadas a cabo para construir la lista de lemas polarizados y mejorados, y, finalmente, se obtuvo un léxico enriquecido para el español que llamaron *iSOL*, éste contiene 8133 palabras clasificadas en dos categorías positivas y negativas.

Para los diversos experimentos se utilizan estos dos lexicones debido a su cobertura en las polaridades positivas, negativas y neutras. Esto quiere decir que tienen una capacidad para caracterizar las posibles etiquetas en las que está categorizado el conjunto de datos del corpus proporcionado. Ambos recursos están disponibles libremente.

Nosotros confiamos en que los recursos *iSOL* y *ML-SentiCon* son recursos valiosos para determinar la polaridad y realizar en análisis de sentimientos para tweets en Español. Adicionalmente, se considera un lexicón propio para uno de los experimentos, obtenido del corpus *InterTASS* y descrito en las siguientes secciones.

Las entradas (frases) de los lexicones son extraídas, pre-procesadas y representadas utilizando el modelo de bolsa de palabras.

## 2.2 Pre-procesado

Esta tarea de pre-procesado de los textos se aplica a los diversos experimentos realizados. Por lo tanto, en esta sección se describen las tareas que involucra el pre-procesado de tweets. Por un lado, para la obtención del lexicón propio y, por otro lado, para las entradas de los lexicones externos *iSOL* y *MLSentiCon*.

La primera tarea para la obtención del lexicón de palabras propio, es la limpieza de los textos. Para ello se realiza una segmentación por palabras (*tokens*) y la eliminación de caracteres especiales, como acentos (*á, é, í, ó, ú*) y signos de puntuación (*. , ¡ ¿ ;*). Después, las unidades léxicas son filtradas eliminando las ligas (*url*) a sitios web externos y las menciones de usuarios en Twitter (*@*). También, se lleva a cabo una normalización de las unidades léxicas

resultantes a minúsculas y se eliminan las *stopwords*, palabras que no aportan significado y por lo tanto, no son funcionales para la identificación de polaridad. Sin embargo, se conservan las palabras de negación (*no, ni*) o afirmación (*si*), al ser consideradas como funcionales para la identificación de la polaridad manifestada por un tweet.

Las tareas de normalización de la risa y la lematización de las palabras son aplicadas a los textos tanto para la obtención de nuestro lexicón como para los experimentos con los lexicones externos que, a priori, se encuentran lematizados.

El objetivo de la normalización de la risa es evitar la redundancia en la forma de expresarla. Para ello, se aplican las reglas o patrones mostrados en la Tabla 1, y se sustituyen las diversas formas de expresar risa por el término en común “jaja”.

Patrón	Frase	Risa normalizada
(ja)+	ja	jaja
(je)+	jeje	jaja
(jo)+	jojojo	jaja
(ji)+	jijijiji	jaja
lol	lol	jaja

Tabla 1: Normalización de la risa

La lematización consiste en obtener la raíz o lema de las palabras, es decir, eliminar los sufijos o flexiones de las palabras. Esto permite agrupar todas las palabras con la misma raíz en una sola representación en el lexicón propio y, mejora el mapeo de los términos de los lexicones externos con los textos de los tweets. Para esta tarea se utiliza la herramienta *TreeTagger* (Schmid, 1999), la cual tiene soporte para el español.

Palabra	Lema
malas	malo
sentirán	sentir
fueron	ir
felices	feliz

Tabla 2: Forma enraizada de palabras para el español

En la Tabla 2, se ha mostrado un ejemplo de palabras en español tal como aparecen en los

textos y con su lema (raíz) que es generada por la herramienta.

### 2.3 Extracción de características y ponderación

La extracción de características se realiza en dos escenarios: para el lexicón propio y para los lexicones externos.

Para el caso de nuestro lexicón, se obtiene la lista de palabras sin repeticiones del conjunto de datos de entrenamiento y desarrollo proporcionado, para la tarea 1 del TASS 2017 denominado corpus *InterTASS* y los corpora generales proporcionados para el TASS 2016. A partir de estos conjuntos se obtienen 9058 palabras. Para los lexicones externos, se extraen sus entradas como unidades y éstas representan las características para la etapa de clasificación del tweet en alguna polaridad determinada.

En ambos escenarios, el del lexicón propio y los lexicones externos, el modelo de bolsa de palabras (*bag-of-words*) es utilizado para la representación de las entradas de los lexicones y la métrica de ponderación denominada *Frecuencia del Término (TF)*. De esta manera, cada tweet, tanto para el conjunto de entrenamiento como para el conjunto de prueba es ponderado con la ocurrencia de cada entrada de los lexicones en un tweet dado.

La ponderación que se ha utilizado, *Frecuencia del Término (TF)*, consiste en el número de veces que un término ( $t$ ) del lexicón aparece en un *tweet (S)*.

De esta manera se obtienen un conjunto de vectores numéricos que representan los tweets del conjunto de entrenamiento y los tweets de los conjuntos de prueba.

### 2.4 Análisis de sentimientos a nivel de tweet.

La tarea de análisis de sentimientos a nivel de tweet, se describe como un tarea típica de clasificación supervisada, la cual consiste en determinar la etiqueta de un tweet basándose en su contenido. Para el caso de la tarea 1 del TASS 2017, se evalúa la clasificación a cuatro niveles de intensidad de polaridad: Positivo “P”, Negativo “N”, Neutro “NEU”, Ninguna “NONE”.

Entonces, para esta tarea de clasificación supervisada se utiliza la representación de los tweets en los vectores resultantes y ponderados con *TF*. Se utiliza el clasificador denominado máquinas de soporte vectorial (SVM) y

presentado en (Chang y Lin, 2001), el cual construye un conjunto de hiperplanos en un espacio  $n$ -dimensional con los tweets de entrenamiento, estos hiperplanos son utilizados para predecir la clase de los tweets de prueba.

El clasificador basado en máquinas de soporte vectorial ha demostrado resultados alentadores en la clasificación de textos. La implementación del algoritmo de clasificación se ha llevado a cabo mediante la herramienta WEKA (Garner, 1995).

Todos los experimentos fueron llevados a cabo con los parámetros: complejidad o número de hiperplanos a construir: -C 1; parámetro gama (tipo de kernel a utilizar): -K PolyKernel; tamaño de la memoria cache a utilizar: -C 250007; parámetro de tolerancia: -L 0.001; y épsilon: -P 1.0E-12.

En los escenarios, de los experimentos, se utilizan los conjuntos de datos de entrenamiento proporcionados por TASS 2016 y TASS 2017 extrayendo los tweets para cuatro etiquetas. La etapa de pruebas del clasificador SVM se lleva a cabo con el conjunto de pruebas del corpus *InterTASS 2017*.

## 3 Experimentos y resultados

El sistema propuesto se evalúa, solamente, para el conjunto de datos de prueba del corpus *InterTASS 2017*, como consecuencia el equipo *LexFAR* solo participa en la tarea 1 del TASS 2017: Análisis de sentimientos a nivel de tweet.

En esta tarea, se realizan tres experimentos utilizando el conjunto de datos de TASS 2016 y TASS 2017 como entrenamiento y prueba, con el corpus *InterTASS* para cuatro etiquetas (P, N, NEU, NONE.).

El primer experimento (run1) consiste en utilizar el lexicón *ML-SentiCon* (run1-msc), para el cual se eliminan aquellas palabras repetidas del lexicón. El resultado es un conjunto de 3084 características que fueron ponderadas para los tweets de entrenamiento y de prueba.

El segundo experimento (run2) consiste en el lexicón propio (run2-propio), el cual está conformado por 9058 entradas o características que fueron ponderadas con la métrica *TF* para los tweets de entrenamiento y prueba.

El tercer experimento (run3) consiste en añadir el lexicón *iSOL* al lexicón *ML-SentiCon* (run3-msc+isol). Al igual que el experimento uno, se eliminan entradas irrelevantes y

repetidas en ambos lexicones. El resultado es un conjunto de 4016 características.

Los tres experimentos fueron desempeñados con los 7184 tweets de entrenamiento y 1899 tweets de prueba para el conjunto de prueba del corpus *InterTASS*, utilizando las máquinas de soporte vectorial como algoritmo de clasificación. Los resultados son presentados en términos de exactitud (Acc) y las versiones macro-promediadas de Precisión (Macro-P), Exhaustividad (Macro-R) y medida F1 (Macro-F1), las cuales serán utilizadas como medidas de evaluación. Los resultados para nuestros tres experimentos, desempeñados para la “Tarea 1: Análisis de sentimientos a nivel de Tweet” del TASS 2017” son mostrados en la Tabla 3.

Experimento	Macro-P	Macro-R	Macro-F1	Acc
run1-msc	0.433	0.427	0.430	0.539
run2-propio	0.372	0.371	0.372	0.490
run3-msc+isol	0.433	0.431	0.432	0.541

Tabla 3: Resultados de nuestros tres experimentos

Los resultados mostrados en la Tabla 3, exponen que nuestro mejor experimento se logra combinando los lexicones *ML-SentiCon* e *iSOL*. Este experimento extiende el número de características ponderadas para cada tweet de 3084 para el primer experimento a 4016 para el tercer experimento. Cabe señalar que la exactitud se logra mejorar de 0.539 para el primer experimento (run1-msc usando el lexicon *ML-SentiCon*) a 0.541 con el tercer experimento (run3-msc+isol usando los lexicones *ML-SentiCon* + *iSOL*).

Los resultados de los diversos equipos o sistemas son clasificados por las medidas Macro-F1 y exactitud (Acc). Por ello, en la Tabla 4, se muestran los resultados del mejor experimento por equipo, para la “Tarea 1: Análisis de sentimientos a nivel de Tweet” del TASS 2017, ordenados de mayor a menor de acuerdo a sus resultados logrados en las medidas Macro-F1 y exactitud. Esto con la finalidad de mostrar la posición de nuestro equipo, LexFAR, con respecto al resto de participantes.

Posición	Equipo	Macro-F1	Acc
1	Jacerong	0.459	0.608
2	ELiRF-UPV	0.493	0.607
3	RETUYT	0.471	0.596
4	tecnolengua	0.441	0.595
5	ITAINNOVA	0.461	0.576
6	SINAI	0.442	0.575
7	<b>LexFAR</b>	<b>0.432</b>	<b>0.541</b>
8	GSI	0.371	0.524
9	INGEOTEC	0.403	0.515
10	OEG	0.377	0.505

Tabla 4: Resultados por equipo de la Tarea 1 del TASS 2017

El mejor experimento por equipo, ordenados por la medida de precisión obtenida, han sido mostrados en la Tabla 4. Con base en estos resultados, nuestro sistema, presentado como LexFAR, ha logrado posicionarse en el lugar 7 con respecto al resto de los participantes. Se ha logrado un valor promedio de F1 de 0.432 y una exactitud de 0.541.

Con la ejecución de los diversos experimentos, se ha detectado que la clase “Ninguna / NONE” ocasiona resultados poco alentadores. Esto se debe, en gran medida, a que las entradas de los lexicones utilizados en nuestro sistema (LexFAR) no aparecen en el conjunto de Tweets de dicha clase.

Los resultados de los tres experimentos son prometedores. El mejor resultado se ha logrado con el experimento denominado run3-msc+isol, el cual utiliza dos lexicones (*ML-SentiCon* e *iSOL*). Sin embargo, es notorio que el experimento 1, que solo utiliza el lexicon *ML-SentiCon*, ha logrado resultados similares al experimento 3. Esto se debe a la similitud de las entradas de los lexicones, ya que *iSOL* se interpone, en gran medida, a *ML-SentiCon*.

Adicionalmente, es notable que el segundo experimento ha quedado por debajo del 0.5 de precisión. A pesar de esto, el sistema propuesto por el equipo LexFAR puede ser utilizado para el análisis de sentimientos a nivel de tweet, con resultados satisfactorios.

#### 4 Conclusiones y trabajo futuro

En este artículo se ha presentado el sistema del equipo LexFAR de la Universidad Autónoma Metropolitana, unidad Azcapotzalco en colaboración con la Benemérita Universidad Autónoma de Puebla para la tarea 1 del TASS 2017: Taller sobre Análisis Semántico en SEPLN. El sistema presentado realiza el análisis de sentimientos a nivel de tweet en español, utilizando lexicones de polaridades semánticas a nivel de lemas para el español. El sistema etiqueta los tweets dependiendo de la intensidad de la polaridad: Positivo (P), Negativo (N), Neutro (NEU) o Ninguno (NONE).

El sistema sometido aborda el análisis de sentimientos como una tarea de clasificación, utilizando el algoritmo de máquinas de soporte vectorial y lexicones para los diversos experimentos enviados. Los lexicones *ML-SentiCon*, *iSOL* y un lexicón propio son utilizados como el conjunto de características que son ponderadas utilizando la métrica de frecuencia de un término.

Nuestro sistema ha logrado posicionarse en el séptimo lugar con respecto al resto de los participantes, considerando sus mejores experimentos. Nuestro mejor experimento se logra combinando los lexicones *ML-SentiCon* e *iSOL*, el cual obtuvo una exactitud de 0.541 para el corpus de prueba del InterTASS. Además, aun cuando el segundo experimento ha quedado por debajo del 0.5 de precisión, el sistema propuesto por el equipo LexFAR puede ser utilizado para el análisis de sentimientos a nivel de tweet, con resultados prometedores.

Como trabajo a futuro nos planteamos experimentar la utilización de otros lexicones, para el español, combinando éstos para la incorporación de nuevas características al sistema de clasificación de tweets en nivel de intensidad de polaridad.

#### Agradecimientos

Este trabajo fue apoyado, parcialmente, por la SEP/PRODEP, con el proyecto DSA/103.5/16/9852. Los autores, también, agradecen a la Universidad Autónoma Metropolitana y a la Benemérita Universidad Autónoma de Puebla.

#### Bibliografía

- Alvarez-López, T., J. Juncal-Martínez, M. F. Gavilanes, E. Costa-Montenegro, F. J. González-Castano, H. Cerezo-Costas, y D. Celix-Salgado. 2015. GTI-Gradient at TASS 2015: A Hybrid Approach for Sentiment Analysis in Twitter. En *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015)*, páginas 35-40.
- Baccianella, Stefano, Andrea Esuli, y Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. En *Proceedings of the Seventh conference on International Language Resources and Evaluation (ELRA)*, páginas 2200-2204.
- Chang, Ch., y Ch. Lin. 2001. LIBSVM - A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3): 27-28.
- Cruz, F. L., J. A. Troyano, B. Pontes, y F. J. Ortega. 2014. ML-SentiCon: Un lexicón multilingüe de polaridades semánticas a nivel de lemas. *Procesamiento del Lenguaje Natural*, 53: 113-120.
- Vilares, D., Y. Doval, M. A. Alonso, C. Gómez-Rodríguez. 2015. LyS at TASS 2015: Deep Learning Experiments for Sentiment Analysis on Spanish Tweets. En *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015)*, páginas 47-52.
- Díaz-Galiano, M. C., y A. Montejo-Ráez. 2015. Participación de SINAI DW2Vec en TASS 2015. En *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015)*, páginas 59-64.
- García-Cumbreras, M. Á., J. Villena-Román, E. M. Cámara, M. C. Díaz-Galiano, M. T. Martín-Valdivia, y L. A. U. López. 2016. Overview of TASS 2016. En *Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference*, páginas 13-21, Salamanca (España).
- Garner, S.R. 1995. Weka: The Waikato environment for knowledge analysis. En *Proceedings of the New Zealand Computer*

- Science Research Students Conference*, páginas 57-64.
- Martínez-Cámara, E., M. Á. García Cumbrreras, M. T. Martín-Valdivia, y L. A. Ureña López. 2015. SINAI-EMMA: Vectores de Palabras para el Análisis de Opiniones en Twitter. En *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015)*, páginas 41-46.
- Martínez-Cámara, E., M. C. Díaz-Galiano, M. Á. García-Cumbrreras, M. García-Vega, y J. Villena-Román. 2017. Overview of TASS 2017, En *Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN (TASS 2017)*, volumen 1896.
- Molina-González, M. D., E. Martínez-Cámara, M. T. Martín-Valdivia, y J. M. Perea-Ortega. 2013. Semantic orientation for polarity classification in Spanish reviews. *Expert Systems with Applications*, 40(18): 7250-7257.
- Montejo-Ráez, A., y M. C. Díaz-Galiano. 2016. Participación de SINAI en TASS 2016. En *Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference*, páginas 41-45.
- Murillo, E. C., y G. M. Raventós. 2016. Evaluación de Modelos de Representación del Texto con Vectores de Dimensión Reducida para Análisis de Sentimiento. En *Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference*, páginas 23-28.
- Quirós, A., I. Segura-Bedmar, y P. Martínez. 2016. LABDA at the 2016 TASS Challenge Task: Using Word Embeddings for the Sentiment Analysis Task. En *Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference*, páginas 29-33.
- Schmid, H. 1999. Improvements in part-of-speech tagging with an application to German. En *Natural language processing using very large corpora*, páginas 13-25. Springer Netherlands.