

A Review of Sentiment Analysis in Spanish

Una Revisión Sobre el Análisis de Sentimientos en Español

Carlos Henríquez Miranda^{1*}, Jaime Guzmán²

¹Universidad Autónoma, Barranquilla, Colombia

²Universitario Nacional de Colombia, Bogotá, Colombia

Received: 11 Dec 2015

Accepted: 6 Sep 2016

Available Online: 7 Dec 2016

Abstract

Sentiment analysis is an area of research that shows an upward trend, especially in the last two years due to large-scale production of opinions and comments from users active over the Web and social networks, in general. Companies and organizations are interested in knowing of their reputation in social networks, blogs, wikis, and other Web sites. Until now, the vast majority of the work involves systems in English. For this reason, the scientific community is interested in studies in other languages. This paper provides a brief overview of the current state of sentiment analysis in Spanish language.

Keywords: Artificial Intelligence, Opinion Mining, Sentiment Analysis, NLP, Emotion Analysis

Resumen

El análisis de sentimiento ha mostrado una alta tendencia de investigación en los últimos dos años debido a la producción a gran escala de opiniones y comentarios por parte de usuarios activos en la Internet. Las empresas y organizaciones, en general, están interesadas en conocer cuál es la reputación que tienen de sus usuarios en las redes sociales, *blogs*, *wikis* y otros sitios Web. Hasta ahora, la gran mayoría de trabajos de investigación involucran sistemas de AS en inglés. Por este motivo, la comunidad científica está interesada en trabajos diferentes a este lenguaje. Este artículo busca dar una breve perspectiva del estado actual del análisis de sentimientos en español.

Palabras clave: Inteligencia Artificial, Minería de Opinión, Análisis de Sentimientos, PLN, Análisis de Emociones

1. Introduction

Currently, the amount of data produced globally is quite high. Companies, governments, universities, and – in general – all organizations produce data at large scale, related to their business. Said data is collected in big repositories, mainly in relational databases that permit structured storage of information. Added to these data, more information is generated daily from the biggest source of all: Internet. It produces millions of data due to the mass use of social networks, messaging services, blogs, wikis, and e-commerce, among others.

This amount of data has required much attention from the scientific community regarding to the production, processing, storage, retrieval, and extraction of information. Just the Web moves millions of non-structured data whose most effective sources are those that offer collaborative environments, like: wikis, social networks, messaging, blogs, and micro blogs among others. This whole range of data is attractive for different commercial, industrial, and academic entities, but extraction and its respective processing makes this task quite complex and difficult if done manually.

*Corresponding Author.

E-mail: chenriquez@uac.edu.co

How to cite: Henríquez C, Guzmán J, *A review of sentiment analysis in spanish*, TECCIENCIA, Vol. 12 No. 22, 35-48, 2017

DOI: <http://dx.doi.org/10.18180/tecciencia.2017.22.5>

To confront this, it is necessary for the extraction, storage and processing of data to be automatic and it is where disciplines, like extraction of information, information retrieval, and natural language processing (NLP) techniques play an important role in managing these large volumes of non-structured data generated daily. The academic community has big work fronts from the vast amount of data from the Web. In it, regular people participate actively in different cloud tools by leaving their comments, opinions and even reviews on all types of themes, using their native language.

Added to this, computers are already starting to acquire the capacity of expressing and recognizing affection, and soon to will have the capacity of “having emotions” [1]. This has been under construction since the emergence of affective computing, which seeks for computers to interpret the emotional state of humans and adapt to their behavior, providing them an adequate response to these emotions. This theme has received attention from researchers in information technology, especially in the field of analysis of emotions where progress has been achieved from the analysis of emotions based on facial expressions [2], recognition of emotions through sensors [3] to the identification of emotions in written texts [4].

The aforementioned would not have been possible without the constant search for new and better models, techniques, tools that allow computers to confront this challenge automatically. For example, one of the current trends in research related to affective computing is sentiment analysis (SA). Sentiment analysis seeks to analyze opinions, sentiments, judgments, attitudes, and emotions of people toward entities, like products, services, organizations, individuals, problems, events, themes, and their attributes [5].

This article sought to show the current state of this topic through a literature review, specifically, progress in the Spanish language. First, a complete review was conducted of SA in general, addressing the basic concepts of the theme. Thereafter, the methodology is discussed and then sentiment analysis in Spanish is addressed by showing the results found. Finally, the conclusions and recommendations are presented.

2. Sentiment analysis or opinion mining

2.1 Definition

In literature, SA receives different denominations or terms. Within these common terms, we find opinion mining, subjectivity analysis, emotion analysis, affective computing, and extraction of the evaluation, among others.

The most-often used in literature are sentiment analysis and opinion mining (OM). According to [6], these are two similar concepts that denote the same field of study, which itself can be considered a sub-field of subjectivity analysis. For [7], these have different origins; OM comes from the information retrieval community whose aim is to extract and elaborate opinions from users about products, films, or other entities.

Sentiment analysis, in turn, was formulated initially as a natural language processing (NLP) task of retrieval of sentiments expressed in texts. Reference [8] states that SA is an area of research in the field of text mining and defines it as the computational treatment of opinions, sentiments, and text subjectivity. Considering the aforementioned, it is noted that most of the terms used are quite similar. Thus, this proposal will address SA differently with OM as an area of NLP work for retrieval of texts, extraction of entities, analysis of opinions, polarity identification (positive or negative), la computational linguistics y all those additional characteristics that permit identifying and extracting subjective information and opinions from textual resources

2.2 Level

Three levels exist to analyze sentiments according to [8] and [9]: at document, phrase, and aspect levels. Analysis at document level classifies the sentiment of a whole document into positive or negative [6]. At phrase level, its objective is to classify the sentiment expressed in each sentence. Sentiment analysis at aspect level seeks to classify sentiment with respect to the specific characteristics of an entity found in each phrase. According to [9], both the document level and phrase level do not discover what it is that people like or do not like, contrary to SA at aspect level, which performs a more profound and detailed analysis. That is, instead of looking at language constructions (documents, paragraphs, sentences, clauses o phrases), SA at aspect level looks directly at the opinion expressed.

2.3 Application

Sentiment analysis is widely used in companies for reputation analysis, that is, how are the organizations positioned in the market according to their clients' opinions in social networks. For this, they use social networks, like Twitter and Facebook as sources to review written texts in form of comments, which contain opinions about their registered brand. This literature review found different types of works framed within different applications; in tourism [10] [11], movie reviews [12] [13], sports [14] politics [15] [16] education [17], health [18], finance [19], and opinion reviews on automobiles [20].

2.4 Steps to perform a sentiment analysis system

According to [6], the goal of an SA system is extraction and classification of the sentiment. Diverse forms of focusing SA exist in literature: [5] [6] [7] [8] [21], and [22]; some are more common than others are. The majority of systems created in literature adopt this series of steps (Figure 1):

1. Extract the information or opinion from a data set,
2. Apply natural language processing techniques, like pre-processing to reduce data noise,
3. Identify the sentiments by locating the characteristics present in the data
4. Classify the sentiment within a polarity scale (positive or negative).

The following describe the last three steps.

2.4.1 Pre-processing

According to [23], pre-processing techniques consist of a text cleaning and preparation process prior to classification. On-line texts generally contain much noise and parts with limited information, like HTML tags, scripts, and notices.

A new trend exists in research on the use of NLP as a preprocessing stage prior to the analysis of sentiments [8]. Different works have been specifically dedicated to this area [23] [24], and [25], which considers that a data preprocessing step is important in sentiment analysis and that with the appropriate selection of techniques, classification precision can be improved.

What takes place in this pre-text processing is, basically, identification of spelling errors, elimination of arbitrary sequences of spaces, stop words, detection of phrase limits, elimination of arbitrary use of punctuation marks, and capitalization among others.

For example, [26] applies different types of pre-processing of NLP in tasks, like: spelling errors, normalization, segmentation, stop words, lemmatization, and name recognition, among others.

2.4.2 Selection of characteristics.

Identifying sentiments is a task of great importance for SA; due to this, many works focus only on identifying sentiments, that is, on selecting characteristics or location in the text of words or phrases that indicate a possible sentiment. For [6], converting a portion of text into a vector of characteristics or another type of representation enables its more outstanding and important traits to be available for data-based systems for text processing. This task comes after data cleaning and seeks to identify where the sentiment is. To address this problem, literature has used distinct approaches [5] [8] [22], like use of terms of presence and frequency, parts of speech (POS), use of opinion words and phrases, and use of denial. In addition, rules [27], syntactic dependence [28], and generic algorithms [29] are used, among others.

2.4.3 Classification of sentiments.

Classification of sentiments (CS) is the task of offering a positive or negative judgment to a comment, opinion, phrase, or document. This task is also known as classification of polarity or as classification of the sentiment of polarity [6]. This is no more than assigning a positive, negative, or neutral value to an opinion. Generally, in literature we find two big approaches for classification of sentiments, according to [7] [8] [30]: based on machine learning and based on lexicon.

For the first approach, a subdivision is made into supervised learning and unsupervised learning. For the lexicon-based approach, it is subdivided based on corpus and based on dictionaries.

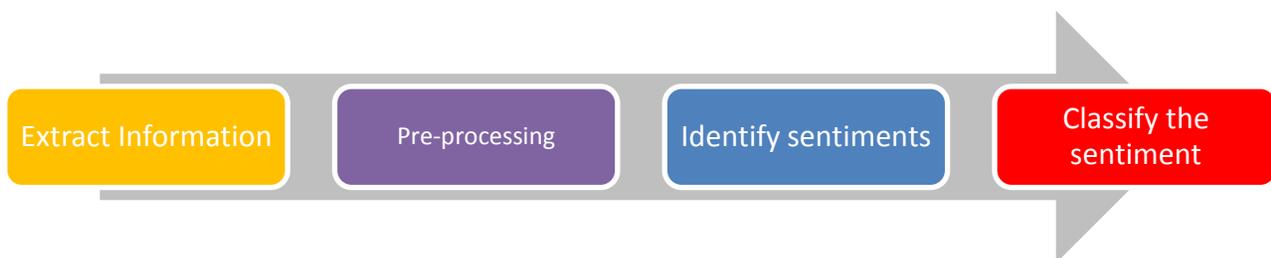


Figure 1 Steps for sentiment analysis

The fundamental differences lie in that the first uses algorithms or strategies to learn from texts or determined corpus and the second model uses dictionaries, lexicons and corpus of words, phrases, or their combination already catalogued with some sentiment.

For a perspective on the research according to the approach used in the literature reviewed, Figure 2 shows the behavior for each of these, finding machine-based learning with 47% and lexicon-based with 32%. This means that works on machine learning are still used frequently in the last six years.

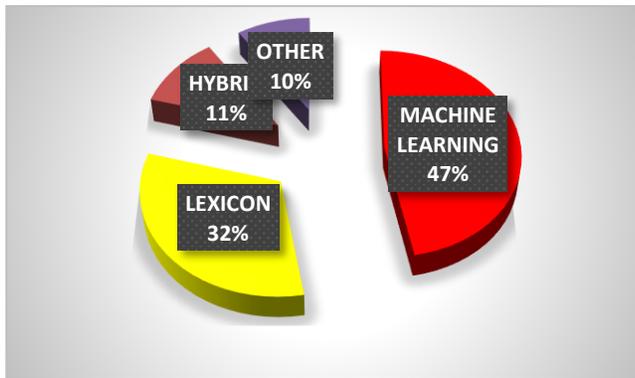


Figure 2 Approaches for classification of sentiments

Many types of classifiers exist in literature based on machine learning, like: those based on support vector machines (SVM) [14] [29] [31] [32] [33] [34] Bayesian classifier [35] [36] [37], neural network [16] [38] and those combining several approaches [16] [38]. With regard to works based on a dictionary or corpus, the following are found: [39] [40] [41] [42] [43] and [44]. In addition, some works are focused on creating, modifying, or improving lexicons for sentiment analysis in different languages [45] [46] [47] [48], and [49].

Of the two approaches presented (Table 1), machine learning techniques have proven extremely useful, not only in the field of sentiment analysis, but also in most text mining and information retrieval applications. In turn, lexicon-based approach is different from approaches based on machine learning in that these are based on lexicon resources generated previously and that store information on the polarity of the elements, which are then identified in the texts and are assigned a polarity. These lexicon-based systems have an advantage over the others by not requiring a training set. Using a given approach will depend on the type of text analysis that will be conducted and if we have training data or sentiment lexicons available. The hybrid approach strategy is also an excellent option.

Table 1 Approaches for classification of sentiments

Approach	Techniques
Machine learning	SVM Naive Bayes Bayesian Network Graphs K-means Neural network Maximum entropy
Lexicon	Dictionary Corpus

2.5 Evaluation metrics for sentiment analysis systems.

Evaluation of these types of systems is conducted from a set of metrics already defined, within which we find the classic performance measurements: *precision*, *recall*, and *F-score*. *Precision* is the number of positive examples classified correctly divided by the number of examples labeled by the system as positive. *Recall* is the number of positive examples classified correctly divided by the number of positive examples in the data [50]. The last serves to correct the distance error in cases where *recall* and *precision* are compensated [51]. Table 2 shows the measurements that depend on confusion matrix in Table 3.

Table 2 Performance measurements for CS systems

Measurement	Formula
Precision	$tp/(tp+fp)$
Recall	$tp/(tp+fn)$
F-score	$(B^2 + 1)tp / ((B^2 + 1)tp + B^2fn + fp)$

Taken from [52]

Table 3 Confusion matrix

Class	Classified as <i>pos</i>	Classified as <i>neg</i>
<i>pos</i>	True positive (tp)	False negative(fn)
<i>neg</i>	False positive (fp)	True negative (tn)

Taken from [52]

3. Methodology

To carry out the review, 32 articles were used focused on themes related to sentiment analysis in the Spanish language. The following base references were chosen: [10] [13] [31] [33] [39] [47] [48] [49] [53] [54] [55] [56] [57] [58] [59] [60] [61] [62] [63] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] and [76].

Table 4 Summary of papers

Article reference	Year	Type of text	Data set	Classification technique	Level of analysis
[61]	2010	AC	WEB	LEX	DOCUMENT
[68]	2010	ACTA	CORPUS	ML	DOCUMENT
[10]	2012	AC	WEB	NA	NA
[53]	2013	ACTA	TWITTER	LEX	DOCUMENT
[13]	2012	AC	CORPUS	HI	DOCUMENT
[33]	2012	ACTA	WORDNET	LEX	NA
[54]	2013	ACTA	TWITTER	ML	DOCUMENT
[47]	2013	ACTA	WEB	ML	DOCUMENT
[55]	2013	AC	TWITTER	HI	DOCUMENT
[56]	2013	AC	TWITTER	OTHER	DOCUMENT
[57]	2013	AC	TWITTER	HI	DOCUMENT
[58]	2013	ACTA	TWITTER	OTHER	DOCUMENT
[59]	2013	ACTA	TWITTER	LEX	DOCUMENT
[60]	2013	ACTA	TWITTER	LEX	DOCUMENT
[62]	2013	ACTA	TWITTER	NA	NA
[39]	2013	AC	TWITTER	LEX	DOCUMENT
[63]	2013	ACTA	YOUTUBE	NA	NA
[65]	2013	AC	TWITTER	LEX	NA
[73]	2013	ACTA	TWITTER	ML	DOCUMENT
[74]	2013	ACTA	TWITTER	ML	DOCUMENT
[49]	2013	AC	CORPUS	LEX	DOCUMENT
[31]	2014	AC	TWITTER	ML	DOCUMENT
[64]	2014	ACTA	TWITTER	ML	ASPECT
[48]	2014	AC	WORDNET	LEX	NA
[66]	2014	AC	TWITTER	ML	DOCUMENT
[67]	2014	ACTA	TWITTER	ML	ASPECT
[69]	2014	ACTA	TWITTER	LEX	PHRASE
[76]	2014	ACTA	NA	NA	ASPECT
[70]	2014	ACTA	WEB	HI	DOCUMENT
[72]	2014	AC	WEB	LEX	DOCUMENT
[75]	2014	ACTA	WEB	LEX	DOCUMENT
[71]	2015	AC	TWITTER	ML	PHRASE

Table 4 illustrates a summary of the articles reviewed. The table has six columns. The first shows the article reference and the second column, the year of publication. The third column shows the type of text as AC, which is a scientific article, and ACTA, which is the result of a scientific event. The fourth column has the data set on which the SA is performed. The following column displays the technique used for the classification, which can be based on LEX (lexicon), ML (machine learning), HI (hybrid), OTHER and NA (not applicable). The last column presents the level of analysis that can be document, aspect, phrase and not applicable.

4. Results and Discussion

Based on the review, some interesting contributions have been found in literature and which are addressed below.

With respect to works on SA under the LEX approach, we found: [49] proposes a new resource for the Spanish research community in sentiment analysis, which generates a new lexicon when translating into Spanish an existing lexicon in English; to evaluate the validity of the lexicon, a series of experiments is proposed. In [61], the authors analyze the opinions of hotel users by employing a system known as *Sentitext* that permits a sentiment analysis system independent of the domain. In [10], although its focus is not computational, analysis is performed of the relations that exist among image, knowledge, brand loyalty, brand quality, and customer value measured through the opinions of tourists in the Trip Advisor virtual community. Reference [39] proposes a method to quantify a user's interest in a theme (at global level) by using SA techniques in Spanish in Twitter. Said work develops a tool called *Tom* that uses a lexicon created semi-automatically by translating from an existing lexicon in English.

With respect to the contributions in SA under ML approach, we find that reference [13] addresses the use of meta-classifiers that combine supervised and unsupervised learning to develop a polarity classification system. This proposal uses a Spanish corpus (*Muchocine*) of movie reviews along with its parallel corpus translated into English. Reference [33] exposes how to combine supervised machine learning algorithms and unsupervised learning techniques for automatic detection of different opinion trends. The proposal has been tested in real textual data available from the comments introduced in a weblog connected to administrative issues in a public education institution.

In addition, the work shows the potential of SA for public organizations and governments to obtain valuable knowledge of opinion trends. Reference [66] presents experiments to study the efficiency of classifying opinions

into five categories: very positive, very negative, positive, negative and neutral by using the combination of the psychological and linguistic characteristics of *LIWC*. *LIWC* is an analysis software that permits extracting different psychological and linguistic characteristics of text in natural language.

Another modality worked in SA in Spanish is to combine LEX and ML techniques. Reference [70] introduces a hybrid approach that uses a lexicon of words tagged according to their polarity, besides automatic learning. The lexicon is generated automatically from a tagged corpus, and a score is assigned to each term from the text for each polarity.

With respect to Table 4, concerning the data source, it is shown that in most works Twitter appears with 60%. This is largely due to the NLP community in Spain [26] that has permitted confronting these new challenges by proposing new challenges for SA in Spanish, like [31] [53] [54] [55] [56] [60] [64] and [67].

With regard to the classification techniques used, it was found that: 42% of the works use lexicon techniques, followed by machine learning with 35%, while hybrid techniques were used in 15% of the works; finally, 8% of the works used another classification technique (Figure 3).

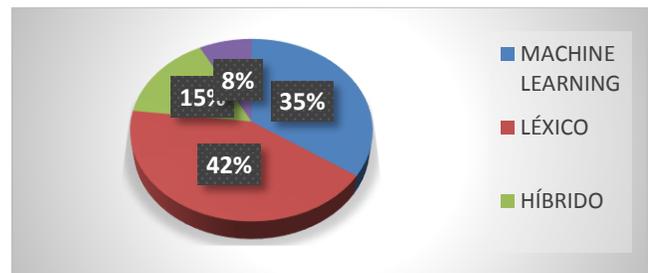


Figure 3 SA works in Spanish by classification techniques

Paradoxically, the works in Spanish use mostly classification techniques with lexicon, although resources and research are still insufficient in this language [8], compared to the rest of the works in other languages, which mostly use machine learning techniques.

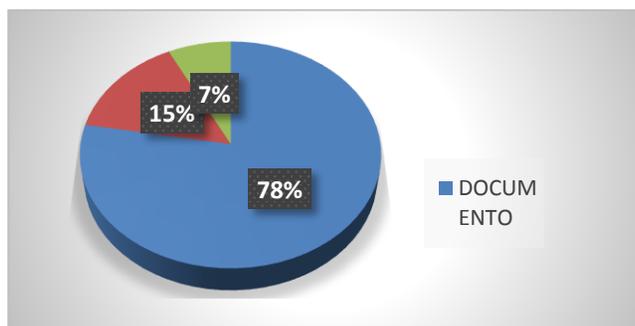
Undoubtedly, linguistic resources used in Spanish for SA tasks are scarce. Some do exist, like *Sentitext*, which has been tested in some SA systems, like [61] and [65], but which is still incomplete because it has some deficiencies in the process of determining polarities, specifically in the middle of the scale (N) and in the very positive (P++) scale. However, some resources exist – as shown in Table 4, which can be a good basis for SA works in Spanish.

Table 4 Lexicon of sentiments used for SA in Spanish

Resource	Author	Year
Sentiwordnet	Esuli, A., & Sebastiani, F. [78]	2006
ANEW	Redondo J. <i>et al.</i> , [79]	2007
Sentitext	Moreno A. <i>et al.</i> , [65]	2010
Emotiblog	Boldrino E. <i>et al.</i> , [34]	2010
ML-Senticon	Cruz F. <i>et al.</i> , [48]	2012

In synthesis, with regard to the classification of sentiments in Spanish, most often traditional techniques are used, like machine learning (ML) and those based on lexicon (LEX). However, LEX techniques tend to lose the battle against ML techniques because they depend largely on quality linguistic resources, especially dictionaries of sentiments [8]. Supervised ML techniques achieve more reasonable efficiency, but the construction of tagging data, that is opinions with sentiment associated, is costly and needs much human work [77]. Contrary to the aforementioned, ML techniques of unsupervised learning do not require tagged training data and can be applicable to other languages and/or domains [27].

With respect to the level of analysis, it was found that most, with 77%, conduct the analysis at document level; the remaining 23% corresponds to 15% at aspect level or characteristic and 8% at phrase level, as noted in Figure 4.

**Figure 4** SA works in Spanish according their level of analysis

According to the prior result, the big difference is seen between the types of analyses performed. Although SA at document level provides a total perspective of that expressed in the opinion, a more detailed analysis is required to extract the most important characteristics of that addressed and find its respective sentiment. It is clear that the quality and rigor of an SA at aspect levels is more complex and difficult.

With regard to the use of SA, it was found that it is widely used in companies to conduct reputation analysis, that is, how the organizations are positioned in the market according the opinions of their customers in social networks. For this, social networks, like Twitter and Facebook were used as sources to review written texts in form of comments that contain opinions about their brand. Besides these, those Web sites where on-line opinions are found, like Trip Advisor and *MuchoCine*, are in good use.

5. Conclusions

This article presents a review in the area of sentiment analysis on contributions made in Spanish since 2012 to 2015. It shows the distinct approaches, levels, data set, and techniques used until now for SA in Spanish.

Initially, it was found that the current state of research on SA in Spanish is scarce compared to other languages, like English. This is probably due to the large amount of linguistic resources existing in this language. From the works reviewed, a large amount of them use these resources in English for their SA systems through automatic translation.

Regarding the techniques used for classification, in SA these are practically distributed between LEX and ML, with a slight advantage of works under the LEX approach, which is somewhat strange due to the scarcity of resources in Spanish.

With respect to the level of analysis, it was found that the vast majority performs it at document level in comparison to aspect level. This is inconvenient when the texts have diverse characteristics with distinct polarities of a product. Hence, new proposals are needed, focused on conducting SA at aspect level, which permit retrieving essential characteristics of a complete document and, thus, perform a more detailed and precise analysis of a product or service.

As a final recommendation, from this article's reflexion, further efforts are encouraged for research on SA in Spanish to potentiate existing linguistic resources. This will permit the construction of new general-purpose models that can analyze different types of written texts with high precision levels.

References

- [1] R. Picard, *Affective Computing*, Cambridge: MIT press, 1997.
- [2] L. Yin, X. Wei and Y. Sun, «A 3D facial expression database for facial behavior research,» from *7th international conference on automatic face and gesture recognition (FGR06)* (pp. 211-216). IEEE., 2006.
- [3] A. Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis and P. F. Driessen, «Gesture-Based Affective Computing on Motion Capture

- Data,» from *International Conference on Affective Computing and Intelligent Interaction*, 2005.
- [4] C. Strapparava and R. Mihalcea, «Learning to Identify Emotions in Text,» from *Proceedings of the 2008 ACM symposium on Applied computing*, 2008.
- [5] B. Liu, *Sentiment Analysis and Opinion Mining*, Synthesis Lectures on Human Language Technologies, 2012.
- [6] B. Pang and L. Lee, «Opinion Mining and Sentiment Analysis,» Vols. %1 of %22(1-2), n° 1-135, 2008.
- [7] M. Tsytsarau and T. Palpanas, «Survey on mining subjective data on the web,» *Data Mining and Knowledge Discovery*, pp. 478-514, 2012.
- [8] W. Medhat, A. Hassan and H. Korashy, «Sentiment analysis algorithms and applications: A survey,» *Ain Shams Engineering Journal*, pp. 1093-1113, 2014.
- [9] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions.*, Cambridge University Press., 2015.
- [10] L. C. Fiol, J. S. García, M. M. T. Miguel and S. F. Coll, «La importancia de las comunidades virtuales para el análisis del valor de marca. El caso de TripAdvisor en Hong Kong y París,» *Papers de turisme*, n° 52, pp. 89-115, 2012.
- [11] C. Henriquez, J. Guzmán and D. Salcedo, «Minería de Opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles,» *Procesamiento del Lenguaje Natural*, vol. 56, pp. 25-32., 2016.
- [12] I. P. Martínez, R. V. García and F. G. Sánchez, «Minería de Opiniones basada en características guiada por,» *Procesamiento del Lenguaje Natural*, n° 46, pp. 91-98, 2011.
- [13] María- Teresa Martín Valdivia, E. M. Cámara, J. M. P. Ortega and L. A. U. López, «Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches,» *Expert Systems with Applications*, vol. 4, n° 10, p. 3934-3942, 2012.
- [14] N. Li and D. D. W. , «Using text mining and sentiment analysis for online forums hotspot detection and forecast,» *Decision Support Systems*, vol. 48, n° 2, pp. 354 - 368, 2010.
- [15] S. Rill, D. Reinelt, J. Scheidt and R. V. Zicari, «PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis,» *Knowledge-Based Systems*, vol. 69, pp. 24-33, 2014.
- [16] M. Anjaria and R. M. R. Guddeti, «A novel sentiment analysis of social networks using supervised learning,» *Social Network Analysis and Mining*, vol. 4, n° 1, pp. 1-15, 2014.
- [17] A. Ortigosa, J. M. Martín and R. M. Carro, «Sentiment analysis in Facebook and its application to e-learning,» *Computers in Human Behavior*, vol. 31, pp. 527-541, 2014.
- [18] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi and L. Donaldson, «Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online,» *Journal of medical Internet research*, vol. 15, n° 11, 2013.
- [19] X. Dong, Q. Zou and Y. Guan, «Set-Similarity joins based semi-supervised sentiment analysis,» *Neural Information Processing*, Springer Berlin Heidelberg, 2012.,» from *Neural Information Processing*, Springer Berlin Heidelberg, 2012, pp. 176-183.
- [20] P. D. Turney, «Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews,» from *Proceedings of the 40th annual meeting on association for computational linguistics*, Stroudsburg, PA, USA, 2002.
- [21] M. Govindarajan and M. Romina, «A Survey of Classification Methods and Applications for Sentiment Analysis,» *The International Journal Of Engineering And Science (IJES)*, vol. 2, n° 12, pp. 11-15, 2013.
- [22] A. Balahur, «Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications,» *Computer Speech & Language*, vol. 28, n° 1, pp. 1-6, 2014.
- [23] L. Dey and S. K. M. Haque, «Opinion mining from noisy text data,» from *Proceedings of the second workshop on Analytics for noisy unstructured text data*, New York, 2008.
- [24] E. Haddi, X. Liu and Y. Shi, «The Role of Text Pre-processing in Sentiment Analysis,» *Procedia Computer Science*, vol. 17, pp. 26-32, 2013.
- [25] J. Smailović, M. Grčar, N. Lavrač and M. Žnidarič, «Stream-based active learning for sentiment analysis in the financial domain,» *Information Sciences*, vol. 285, pp. 181-203, 2014.
- [26] Sepln, 12 12 2014. [Online]. Available: <http://www.sepln.org/>.
- [27] A. Bagheri, M. Sarace and F. d. Jong, «An Unsupervised Aspect Detection Model for Sentiment Analysis of Reviews,» from *Language Processing and Information Systems*, Springer Berlin Heidelberg, 2013, pp. 140-151.
- [28] G. Qiu, B. Liu, J. Bu and C. Chen, «Opinion word expansion and target extraction through double propagation,» *Computational Linguistics*, vol. 37, n° 1, pp. 9 - 27, 2011.
- [29] A. Abbasi, H. Chen and A. Salem, «Sentiment Analysis in Multiple Languages: Feature,» *ACM Transactions on Information Systems (TOIS)*, vol. 26, n° 3, p. 12, 2008.
- [30] K. Ravia and V. Ravia, «A survey on opinion mining and sentiment analysis: Tasks, approaches and applications,» *Knowledge-Based Systems*, vol. 89, pp. 14-46., 2015.
- [31] F. Pla and L.-F. Hurtado, «Sentiment Analysis in Twitter for Spanish,» *Natural Language Processing and Information Systems*, pp. 208-213, 2014.
- [32] T. Mullen and N. Collier, «Sentiment Analysis using Support Vector Machines with Diverse Information Sources,» *EMNLP*, vol. 4, pp. 412-418, 2004.
- [33] César Alfaro, «A multi-stage method for content classification and opinion mining on weblog comments,» *Annals of Operations Research*, pp. 1-17, 2013.
- [34] E. Boldrini, A. Balahur, P. Martínez-Barco and A. Montoyo, «Using EmotiBlog to annotate and analyse subjectivity in the new textual genres,» *Data Mining and Knowledge Discovery*, vol. 25, n° 3, pp. 603-634, 2012.
- [35] A. Lazaridou, I. Titov and C. Sporleder, «A Bayesian Model for Joint Unsupervised Induction of Sentiment, Aspect and Discourse Representations,» *ACL*, vol. 1, pp. 1630 -1639, 2013.
- [36] A. Pak and P. Paroubek., «Twitter as a Corpus for Sentiment Analysis and Opinion Mining,» *LREC*, vol. 10, pp. 1320 -1326, 2010.
- [37] J. Ortigosa-Hernández, J. D. Rodríguez, L. Alzate, M. Lucania, I. Inza and J. A. Lozano, «Approaching Sentiment Analysis by using semi-supervised learning of multi-dimensional classifiers,» *Neurocomputing*, vol. 92, pp. 98-115, 2012.
- [38] M. Ghiassi, J. Skinner and D. Zimbra, «Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network,» *Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network*, vol. 40, n° 16, pp. 6266-6282, 2013.
- [39] F. Rodríguez, Cuantificación del interés de un usuario en un tema mediante minería de texto y análisis de sentimiento, Diss. Universidad Autónoma de Nuevo León, 2013.
- [40] T. Nasukawa and J. Yi, «Sentiment analysis: Capturing favorability using natural language processing,» from *Proceedings of the 2nd international conference on Knowledge capture. ACM*, 2003.
- [41] J. Brooke, M. Tofiloski and M. Taboada, «Cross-Linguistic Sentiment Analysis: From English to Spanish,» *RANLP*, pp. 50 - 54, 2009.

- [42] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, «Lexicon-Based Methods for Sentiment Analysis,» *Computational linguistics*, vol. 37, n° 2, pp. 267 - 307, 2011.
- [43] Alexander Hogenboom, «Multi-lingual support for lexicon-based sentiment analysis guided by semantics,» *decision support systems*, vol. 62, pp. 43-53, 2014.
- [44] S. Feng, D. Wang, G. Yu, W. Gao and K.-F. Wong, «Extracting common emotions from blogs based on fine-grained sentiment clustering,» *Knowledge and information systems*, vol. 27, n° 2, pp. 281-302, 2011.
- [45] F. Cruz, J. T. J. Ortega and C. Vallejo, «Inducción de un Lexicón de Opinión Orientado al Dominio,» *Procesamiento del lenguaje natural*, vol. 43, pp. 5 - 12, 2009.
- [46] S. Baccianella, A. Esuli and F. Sebastiani, «SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining,» *LREC*, vol. 10, pp. 2200 -2204, 2010.
- [47] V. Perez-Rosas, C. Banea and R. Mihalcea, «Learning Sentiment Lexicons in Spanish,» *LREC*, pp. 3077-3081, 2012.
- [48] F. L. Cruz, J. A. Troyano, B. Pontes and F. J. Ortega, «Building layered, multilingual sentiment lexicons at synset and lemma levels,» *Expert Systems with Applications*, vol. 41, n° 13, pp. 5984-5994, 2014.
- [49] M. D. Molina-González, E. Martínez-Cámara, M.-T. Martín-Valdivia and J. M. Perea-Ortega, «Semantic orientation for polarity classification in Spanish reviews,» *Expert Systems with Applications*, vol. 40, n° 18, pp. 7250-7257, 2013.
- [50] M. Sokolova and G. Lapalme, «A systematic analysis of performance measures for classification tasks,» *Information Processing and Management*, pp. 427 - 437, 2009.
- [51] M. John, F. Kubala, R. Schwartz and W. R., «Performance measures for information extraction,» from *Proceedings of DARPA broadcast news workshop*, 1999.
- [52] M. Sokolova and G. Lapalme, «A systematic analysis of performance measures for classification tasks,» *Information Processing and Management*, pp. 427 - 437, 2009.
- [53] E. Rodríguez, A. I. T. Bastida, A. García-Serrano and M. G. Rodríguez, «Using a linguistic approach for sentiment analysis,» from *TASS 2011*, 2011.
- [54] Pablo Gamallo, M. G. and S. Fernández-Lanza, «TASS: A Naive-Bayes strategy for sentiment analysis on Spanish tweets,» from *Workshop on Sentiment Analysis at SEPLN (TASS2013)*, 2013.
- [55] J. Fernandez, Y. Gutiérrez, J. M. Gómez, P. Martinez, A. Montoyo and R. Muñoz, «Sentiment Analysis of Spanish Tweets,» from *XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural (SEPLN 2013)*, 2013.
- [56] H. Cordobés, «Técnicas basadas en grafos para la categorización de tweets por tema,» from *XXIX Congreso de la Sociedad Española de Procesamiento del Lenguaje Natural (SEPLN 2013)*, 2013.
- [57] X. Saralegi Urizar and I. San Vicente Roncal, «Elhuyar at TASS 2013,» from *TASS 2013*, Madrid, 2013.
- [58] R. del Hoyo, I. Hupont and F. Lacueva, «Descubrimiento de nuevas palabras con polaridad afectiva através de técnicas de inteligencia artificial general,» from *TASS 2013*, Madrid, 2013.
- [59] B. Alexandra and J. M. Perea-Ortega, «Experiments using varying sizes and machine translated data for sentiment analysis in Twitter,» from *TASS workshop at SEPLN 2013*, 2013.
- [60] D. Garcia and M. Thelwall, «Political alignment and emotional expression in Spanish Tweets,» from *XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural*, Madrid, 2013.
- [61] A. Moreno, F. P. Castillo and R. H. García, «Análisis de Valoraciones de Usuario de Hoteles con Sentitext,» *Procesamiento del Lenguaje Natural*, vol. 45, pp. 31-39, 2010.
- [62] I. Alegria, N. Aranberri, V. Fresno, P. Gamallo, L. Padró, I. San Vicente, J. Turmo and A. Zubiaga, «Introducción a la tarea compartida Tweet-Norm 2013: Normalización léxica de tuits en español,» from *XXIX Congreso de la Sociedad Española para el Procesamiento del Lenguaje*, Madrid, 2013.
- [63] V. Rosas, R. Mihalcea and L.-P. Morency, «Multimodal sentiment analysis of Spanish online videos,» *IEEE Intelligent Systems*, vol. 28, n° 3, pp. 38 - 45, 2013.
- [64] L. Hurtado and F. Pla, «Análisis de Sentimientos, Detección de Tópicos y Análisis de Sentimientos de Aspectos en Twitter,» from *TASS 2014*, 2014.
- [65] A. Moreno-Ortiz and C. P. Hernández, «Lexicon-based sentiment analysis of Twitter messages in Spanish,» *Procesamiento del lenguaje natural*, vol. 50, pp. 93-100, 2013.
- [66] M. Salas-Zárate, E. López-López, R. Valencia-García, N. Aussenac-Gilles, Á. Almela and G. Alor-Hernández, «A study on LIWC categories for,» *Journal of Information Science*, vol. 40, n° 6, pp. 749-760, 2014.
- [67] D. Vilares, Y. Doval, M. A. Alonso and C. Gómez-Rodríguez, «A Prototype for Extracting and Analysing Aspects from Spanish tweets,» from *TASS 2014*, 2014.
- [68] A. Rosá, D. Wonsever and J.-L. Minel, «Opinion Identification in Spanish Texts,» from *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*, Los Angeles, 2010.
- [69] D. Vilares, M. Hermo, M. A. Alonso, C. Gomez-Rodríguez and Y. Doval, «LyS: Porting a Twitter Sentiment Analysis Approach from Spanish to English,» from *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, Ireland, 2014.
- [70] J. Fernández, J. M. Gómez and P. Martínez-Barco, «Análisis de sentimientos multilingüe en la Web 2.0,» from *V Jornadas TIMM*, Cazalla de la Sierra, 2014.
- [71] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Ureña-López and R. Mitkov, «Polarity classification for Spanish tweets using the COST corpus,» *Journal of Information Science*, 2015.
- [72] M. D. Molina-González, E. Martínez-Cámara, M. T. Martín-Valdivia and L. A. Ureña-López, «A Spanish semantic orientation approach to domain adaptation for polarity classification,» *Information Processing & Management*, 2014.
- [73] J. Fernández, Y. Gutiérrez, J. M. Gómez, P. Martínez-Barco, A. Montoyo and R. Muñoz, «Análisis de sentimientos sobre tweets en castellano utilizando un algoritmo de ranking y skipgrams,» from *TASS 2013*, Madrid, 2013.
- [74] A. Montejo-Ráez, M. C. Díaz-Galiano, J. M. Perea-Ortega and L. A. Ureña-López, «Spanish knowledge-based generation for polarity classification from masses,» from *Proceedings of the 22nd international conference on World Wide Web companion*, Ginebra, 2013.
- [75] M. D. Molina-González, E. Martínez-Cámara, M. T. Martín-Valdivia and L. A. Ureña-López, «Cross-Domain Sentiment Analysis Using Spanish Opinionated Words,» from *19th International Conference on Applications of Natural Language to Information Systems*, Montpellier, 2014.
- [76] J. Villena Román, J. García Morera, E. Martínez Cámara and S. M. Jiménez Zafra, «TASS 2014 - The Challenge of Aspect-based Sentiment Analysis,» from *TASS 2014*, Madrid, 2014.
- [77] S. Brody and N. Elhadad, «An unsupervised aspect-sentiment model for online reviews,» from *uman Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 2010.
- [78] A. Esuli and F. Sebastiani, «Sentiwordnet: A publicly available lexical resource for opinion mining,» from *In Proceedings of LREC*, 2006.

[79] J. Redondo, I. Fraga, I. Padrón and M. Comesaña, «The Spanish adaptation of ANEW,» *Behavior Research Methods*, vol. 3, p. 39, 2007.