ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 11(22), DOI: 10.17485/ijst/2018/v11i22/102251, June 2018*

Is News Really Pessimistic? Sentiment Analysis of Chilean Online Newspaper Headlines

Claudia Martinez-Araneda[1]*, Alejandra Segura[2], Christian Vidal-Castro[2] and Jorge Elgueta[2]

[1]Computer Science Department, Universidad Católica de la Santísima Concepción. Alonso de Ribera 2850, Concepción, Chile; cmartinez@ucsc.cl

[2]Information System Department, Universidad del Bío-Bío, Av. Casilla 5-C, Collao 1202, Concepción, Chile; asegura@ubiobio.cl, cvidal@ubiobio.cl, jorelgue@alumnos.ubiobio.cl

Abstract

Objectives: This paper explores the popular belief that all news is bad news. Many claim not to read newspapers to avoid knowing about the worst of our society. We want tear down the myth by applying a Sentiment Analysis (SA) approach. **Method/Analysis:** This work applies sentiment analysis techniques to study the headline bias of online newspapers for the period between March 2014 and April 2015. We analyzed 2953 headlines gathered from five of the most popular Chilean newspapers which are available online and offer RSS feeds. **Findings**: Our results show a roughly equivalent percentage of positive bias (38%) and negative bias (37%) instances, with 25% of headlines exhibiting a neutral bias. Automatic classification performance is promising, with decent classifier performance and sensitivity, with plenty of room for improvement. **Novelty/Improvement:** This work also a domain-specific Spanish language tagged corpus was generated as a result of this work, which is a valuable resource for future studies.

Keywords: Bias, Sentiment Analysis, Subjectivity Analysis, Text Mining

1. Introduction

This work uses the lexicon approach to detect bias in the headlines of five Chilean newspapers available online and that offer RSS feeds. The analysis uses the unified lexicon proposed in[1] which was generated by merging the full-strength lexicon by[2] and the lexicon proposed in[3]. Using this unified lexicon, a bag-of-words approach was used to analyze bias in headlines published between March of 2014 and April of 2015, on the basis of 4862 words tagged as positive, negative, or neutral.

The remainder of the article is organized as follows: the next section presents a bibliographic review of articles related to sentiment analysis, with an emphasis on lexicon-based approaches. Section3 describes the methodology, including descriptions of the techniques and tools used in the information extraction and in the preprocessing phases, and the automatic classification algorithm. This section also describes the experiments used to evaluate the automatic classifier's performance using experts' opinions. Analysis and discussion of results is shown in section4. Finally, section5 presents our conclusions and outlines future work.

Sentiment Analysis (SA), a term first coined by[4,5], is a branch of Natural Language Processing (NLP) aimed at the identification and/or classification of bias in a written text. It has also been called opinion mining by Dave et al.[6] and subjectivity analysis by Wiebe et al[7]. SA encompasses various tasks such as subjectivity detection, bias identification, opinion detection and the detection of emotions and their intensity. Depending on framework in which is based the SA is also called SA (or opinion mining) aspect-based or SA (or opinion mining) feature-based[8,9]. In the broadest sense this field of research has also been called sentic computing[10]. Another definition of SA is assign to the task of identifying opinions, emotions, ratings, using computer processing power to formalize and polarize this content[11].

*Author for correspondence

The information described in[12] is of great interest where three elements of study are included. These are the polarity, the strength and the emotion defined as:

- Polarity: These methods and resources allow obtaining polarity from a text.
- Strength: These methods include intensity levels according to a detected polarity.
- Emotion: These methods and resources are focused on extracting emotion or mood states from a text based on some taxonomy of emotions from the psychology area.

Sentiment analysis is generally performed by using either machine learning techniques, lexical resources or hybrid approach. The machine learning methods usually apply supervised or unsupervised learning algorithms to the detection and classification of texts[13,14].

Alternatively, bias is detected by using lexical structures that contain sets of previously tagged words[3-15]. Hybrid approaches that are interesting are those applied by[16,17] among others.

In general terms, the machine learning approach requires a large volume of tagged data; both for the training of the learning algorithms and for their evaluation, and results can be affected by the corpus. The lexicon approach requires less data, but is heavily dependent on the completeness and correctness of the lexical resources used.

According[18] most lexicon-approaches are based on English language. If we consider that the Spanish language is the third most used language on the web after the English and Chinese languages and that the web users that communicate in Spanish have grown by 807.4% between 2000 and 2011 and also that Spanish is the second most used language in the two main social networks in the world: Facebook and Twitter[19]. For the above reasons, we can say that the development of lexical resources and tools for subjectivity and SA in languages other than English is a growing need.

In recent years social networks have been incorporated as a source of data for SA in answering to penetration of these applications in Internet users (Figure 1) according to the data obtained from[15]. It is interesting the processing the continuous flow of data from twitter where often it is necessary incorporate an analysis of subjectivity in real time. To apply sentiment analysis and opinion mining with this large amount of data some streaming techniques[20,21] should be considered in order to obtain results quickly. Some interesting works include one that describes a system for real time analysis of public sentiment toward presidential candidates in[21], and another contains an approach for political tendency identification where some metrics are defined that take into account the polarity of political entities in the tweets of each user[22].

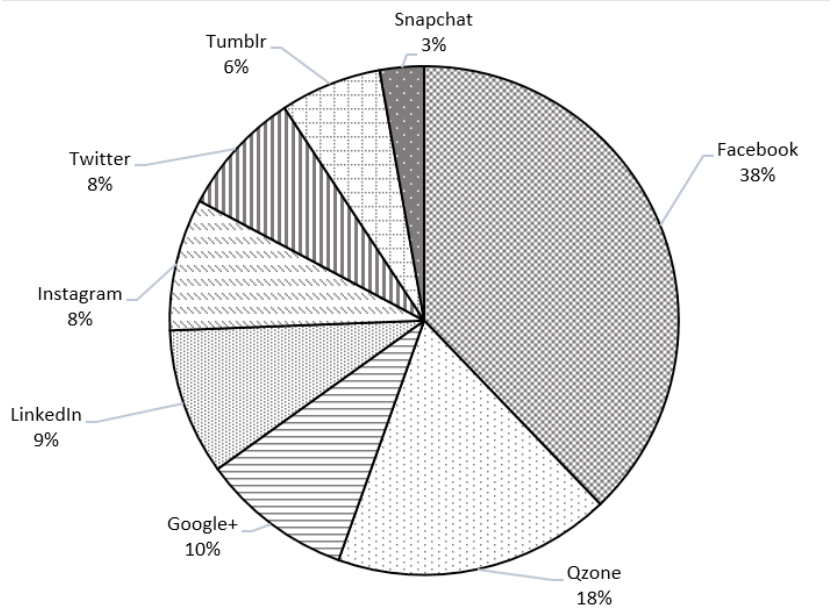

Figure 1. Active users in social network December 2014 (Source: Facebook, Google, Twitter, WhatsApp).

According to Purita[23] the growth trend of users of social networks is estimated to double by 2018 in comparison to the amount of users in 2011 (2440 million vs. 1220).

Finally, the domains are varied in which ones are needed to detect or predict the opinions of Internet users in relation to social issues such as healthcare, politics, government, finance, education and judicial power, among others.

2. Methodology

As is shown in Figure 2, the methodology can be divided into five main phases: headlines extraction, preprocessing, corpus generation, automatic classification and classifier performance evaluation. As is show in Figure 2, some stages were performed in parallel way.

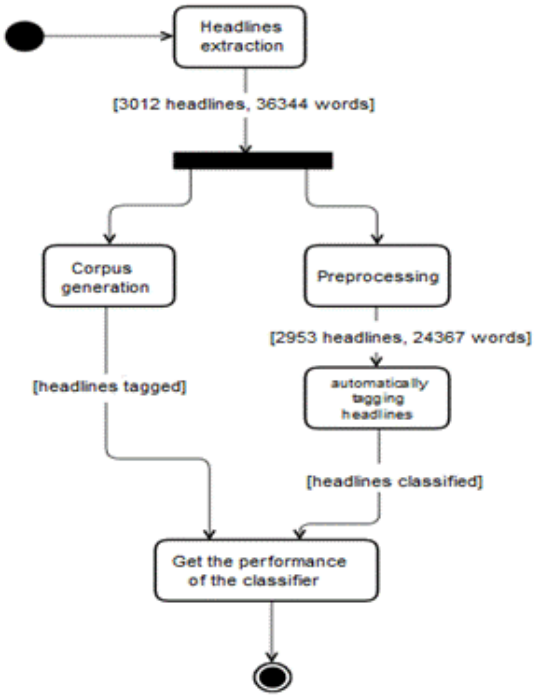

Figure 2. Methodology workflow.

2.1 Headlines Extraction

We did an exploratory search of Chilean newspapers which were available online and offered RSS feeds. Headlines were extracted from the El Mercurio, El Mostrador, La Tercera, Soy Chile and Terra newspapers. We excluded newspapers lacking RSS feeds and those which, because of their sarcastic and/or ironic nature, would require special treatment and distort the results. An initial set of 3012 headlines and 36344 words was generated as a result of this phase.

2.2 Preprocessing

Before the classification phase, the texts must be normalized, cleansed and prepared. Preprocessing helps improve classifier performance and speeds up the classification process, allowing real-time affective analysis. It involves several steps: data cleansing, whitespace removal, abbreviation expansion, stemming, and negative phrase handling and filtering[24]. Figure3 illustrates the sequence of steps involved in headline preprocessing.

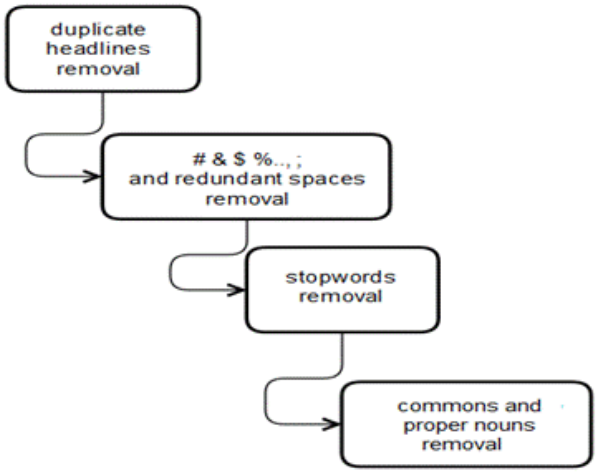

Figure 3. Headline preprocessing.

The main tool used in this phase is R-Project version 3.0.1[25] and its packages SnowballC, gsubfn, RCurl, Rjson y Tm, which are aimed at data preprocessing and corpus generation. After the preprocessing phase, and given that most proper and common nouns, by themselves, do not define a sentence's bias, we chose to remove proper nouns using a 558-word Spanish-language list, and common nouns by using a 1007-word Spanish-language list. Stemming was performed only for those words not present in the lexicon.

As can be seen in Table 1, corpus characterization after preprocessing for all five newspapers gives us a total of 2953 headline instances.

Table 1. Corpus characterization

Newspaper	URL source	Headlines	Words
El Mercurio	http://www.emol.cl	1138	9978
El Mostrador	http://www.elmostrador.cl	342	3085
La Tercera	http://www.latercera.cl	296	2385
Soy Chile	http://www.soychile.cl	237	1937
Terra	http://www.terra.cl	940	6982
Σ		2953	24367

Figure 4 presents a tag cloud representing a snapshot of the Chilean news context for the period between March 2014 and April 2015, containing only higher-frequency words (>50). This word frequency visualization can be explained by noting that during this period the new Chilean tax law (law number 20780) was in its parliamentary discussion period, and it was finally signed into law in September 2014. Thus, this law and its related topics became trending topics in social media networks from April to September 2014.

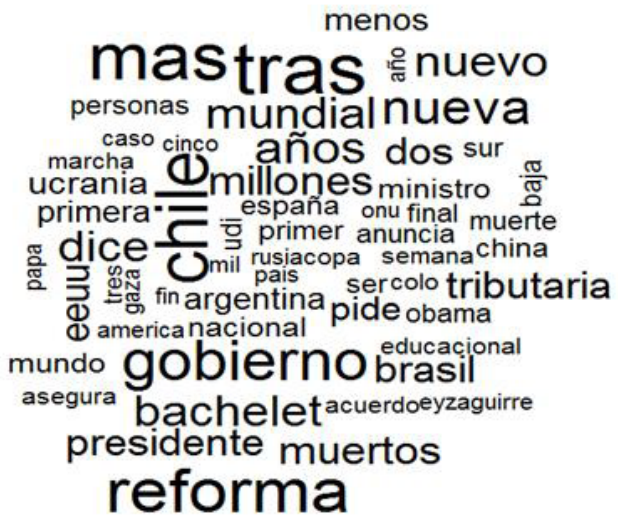

Figure 4. Tag Cloud headlines.

2.3 Automatic Classification

This work uses the unified lexicon proposed in[1], which is generated by merging the full-strength lexicon of[2] and the lexicon proposed in[26]. The resulting lexicon has 4862 entries tagged as positive, negative or neutral. To determine a headline's bias, it was represented as a bag-of-words, and each word was compared to the lexicon's tagged words, as described in the following algorithm (Figure 5).

```
for j=1, size(headlines)
for each word in headlines[j]
    for k=1, size(lexicon)
    if (lexicon[k].word = word)
Labeled (headlines[j], word, lexicon[k].polarity)
    else
        if (stem(lexicon[k].word) = stem(word))
        Labeled (headlines[j], lexicon[k].polarity)
        else
        Labeled (headlines[j],"not found")
        Endif
    Endif
Next k
Next word
Next j
```

Figure 5. Own match algorithm.

Finally, we applied (1) as variant of Emotion Word Density metric[27] to calculate each headline's bias. For each headline *h*, all positive (+1), negative (-1) and neutral (0) instances were added and the final headline bias calculated by applying the following rules:

if bias(h) > 0, then h is positive
if bias(h) = 0, then h is neutral
if bias(h) < 0, then h is negative

$$\text{Bias(h)} = \sum_{1}^{k1} \text{instances (+) (h)} - \sum_{1}^{k2} \text{instances (-) (h)}$$

where, bias (h) is the bias of each headline h considering $k1$ positive instances y $K2$ negative instances and $k1$, $K2$ ≥0. A fourth case, No bias (N/B) was assigned to all headlines that exhibited no coincidences between its constituent words and the lexicon. These should not be confused with neutral headlines, that is, a headline for which mathematically it obtained a 0 in the value of polarity (1).

As can be seen in Table 2, results show that only 43.9% of all headlines have a negative, positive or neutral bias, and that 56.1% of all headlines were tagged as No Bias (N/B).

Table 2. Automatic classification results

Percentage		Bias
21.3%	43.9%	Negative
18.3%		Positive
4.3%		Neutral

2.4 Classifier Performance Evaluation

The classifier's performance was evaluated by comparing its performance against manual tagging by experts, using traditional information retrieval metrics. Manual tagging was done via a special-purpose tool (Figure 6), which allowed experts to assign bias to each headline, this tool also allowed us tagging the newspapers headlines based on emotion classification of Plutchik for the future research in the area of affect analysis. Experts were chosen so they belong to the same age group, have similar abilities with IT and are frequent readers of news online. All of them were trained in the tagging protocol, which emphasizes the difference between biased headlines and headlines that state facts and thus are not, by themselves, biased

	Headline	rater 1	rater 2	rater 3	polarity
1	banco francia preve crecimiento segundo trimestre año	positivo	positivo	positivo	positivo
2	joven guardia seguridad muere tras ser acribillado pudahuel	negativo	negativo	negativo	negativo
3	britney spears extiende temporada show vegas	neutro	positivo	positivo	positivo
4	rusia negociara rebajas gas ucrania pague deudas	negativo	negativo	negativo	negativo
5	alemania coloca bonos us millones seis meses	positivo	neutro	neutro	neutro
6	presidenta bachelet reune lunes primera visita	neutro	positivo	positivo	positivo
7	cervecera española mahou san firma acuerdo empresa filipina	positivo	ıtro	positivo	positivo
8	trafico pasajeros air france klm crecio abril		negativo	positivo	x
9	piden ex take that devuelva condecoracion ii tras evadir impuestos		positivo	negativo	negativo
10	bolsa tokio cierra baja inestabilidad ucrania subida yen		negativo	negativo	negativo
11	psiquiatra asegura pistorius padecia trastorno ansiedad	positivo	negativo	negativo	negativo
12	expertos seguridad google alertan vulnerabilidad medios	negativo	negativo	negativo	negativo
13	encuesta banco central preve mantencion tasa politica monetaria	neutro	neutro	neutro	neutro
14	mundo rock manchester tambien celebro titulo city	positivo	positivo	positivo	positivo
15	dolar abre sesion baja presionado importante alza cobre londres	positivo	neutro	neutro	neutro

negativo
neutro
positivo
rater 1

Figure 6. Manual tagging tool (example).

These experts classified the headlines as having positive, negative or neutral bias. The agreement level was calculated using two multi-rater metrics (Krippendorff's alpha y Cohen's Kappa) and also using the ReCal web service[28], obtaining a moderate degree of agreement (0.466). Starting with the experts' evaluation, the most frequent bias was chosen using single bias assignment criteria: if more than 2/3 of all experts agree on a headline´s bias, then that bias is assigned to the headline. On the other hand, if the experts are in disagreement, then that headline was eliminated. This only occurred in less than 3% of all cases. In this manner, the corpus was whittled to 2859 headlines and 23714 words.

Figure 7 presents the results of the manual tagging process. It can be seen that the bias of the corpus is 38% positive, 37% negative and 25% neutral bias. Likewise, Figure8 shows that, according to the experts, the newspapers with the highest number of positive bias headlines are Soy Chile and Emol, while the newspaper with the highest number of negative bias headlines is El Mostrador.

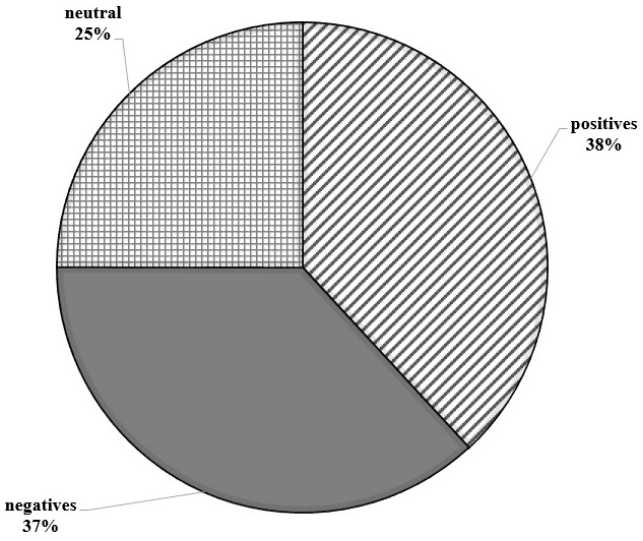

Figure 7. Corpus bias distribution.

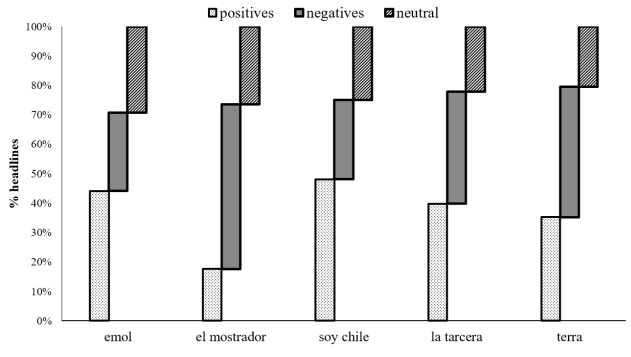

Figure 8. Headline bias distribution per newspaper.

Figure 8 also shows that negative bias headlines are predominant in El Mostrador and Terra, while the Soy Chile, Emol and La Tercera newspapers have a majority of positive bias headlines. This stands in contrast to the rather homogeneous distribution of positive bias and negative bias headlines in the corpus.

Once the tagged corpus and classifier results were available, the classifier's performance was evaluated using the standard information retrieval metrics. As mentioned in Table 2, 1257 headlines (43.9%) were used as the basis for calculations. Table 3 shows the results calculated for the accuracy, precision, recall and f-measure metrics. This resource will serve to a new proposal based on the statements "all emotions vary in their degree of similarity to one another", "each emotion can exist in varying degrees of intensity"[29], since our idea is to incorporate similarity metrics to improve analysis.

Table 3. Classifier performance

Measure	Performance
Accuracy	0.41209
Precision	0.43048
Recall	0.33934
F-measure	0.37951

3. Results and Discussion

Comparing the results from the automatic classifier to the results from the tagged corpus generation process shows moderate conformity with the real values across the set (accuracy = 41%), a regular predictive capacity (precision = 43%), and a sensitivity that can be improved (recall = 33%). From comparing the results in the evaluation phase, we have identified 3 situations that require further analysis:

- The classifier tags a headline as not having bias (N/B), while the experts tag it as having positive, negative or neutral bias. As was shown in Table 2, 56.1% of all corpus headlines were tagged as not having bias by the classifier. 73% of these headlines were tagged as having a positive or negative bias by the experts, as shown in Table 4a. This situation emphasizes the importance of working with complete and correct lexical resources.
- The classifier tags a headline as not having bias (N/B), while the experts tag it as being a factual headline (Table 4a).
- The classifier tags a headline as having a positive or negative bias, while the experts tag it as being a factual headline (Table 4b).

Table 4a. Automatic classification v/s experts

Automatic classification	Experts		
N/B (56% of the corpus)	Positive	Negative	Fact
	38%	35%	27%
	73%		

Table 4b. Automatic classification v/s experts

Automatic classification	Experts
Positive or negative	Fact
	8%

4. Conclusion and Future Work

This work helped us find out the headline bias of Chilean online newspapers and question the belief that all news are bad news, that is to say, that news headlines throughout the media landscape (radio, TV, blogs, web, etc.) are predominantly negative. Our results show that positive headlines and negative headlines are evenly distributed.

Regarding the scenarios presented in the discussion section, we can identify possible improvements and future work: in the case of scenario a) all words present in the headlines tagged by the experts as having a bias can be added to the lexicon. In the case of scenario c) factual headlines can be recognized in the preprocessing phase by, for example, adding a headline grammar analysis step. Both approaches would help reduce the number of headlines tagged as N/B and improve the automatic classifier.

An important contribution of this work is the generation of a Spanish-language domain-specific tagged corpus comprised of online news headlines, which can be used

as a lexical resource for other studies. In this work we considered too the labeled of headlines using the model of eight emotions of Plutchik on the labeled corpus. This resource will serve to a new proposal based on the statements "all emotions vary in their degree of similarity to one another", "each emotion can exist in varying degrees of intensity"[29], since our idea is to incorporate similarity metrics to improve analysis in the future.

As future work also, we hope to apply a framework for irony detection and compare results.

5. Acknowledgement

This work has been done in collaboration with the research group SOMOS (SOftware-MOdelling-Science) funded by the Research Agency and the Graduate School of Management of the Bío-Bío University under grant 130415 GI/EF and the Engineering Faculty and Computer Science Department of the Universidad Católica de la Santísima Concepción, Chile.

6. References

1. Oyarzún C, Segura A, Vidal-Castro C, Martinez-Araneda C, Rubio-Manzano C. Análisis automático de sentimientos sobre comentarios de novelas en espa-ol. Contributions to the uses of Technologies for Learning. CCita. Miami: Humboldt International University; 2015. p. 316–23.
2. Pérez-Rosas V, Banea C, Mihalcea R. Learning Sentiment Lexicons in Spanish; 2011. p. 1–5.
3. Strapparava C, Mihalcea R. Learning to identify emotions in text. Proceedings of the ACM Symposium on Applied Computing. New York, NY, USA: ACM; 2008. p. 1556–60.
4. Turney P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia, Pennsylvania, USA for Information Technology Institut / National Research Council of Canada; 2002. p. 417–24.
5. Pang B, Lee L, Rd H, Jose S. Thumbs up? Sentiment Classification using Machine Learning Techniques; 1988. p. 1–8.
6. Dave K, Lawrence S, Pennock MD. Mining the peanut gallery: Opinion extraction and semantic classification of product. Proceedings of the 12th international conference on World Wide Web; 2003. p. 519–28. Crossref.
7. Wiebe J, Wilson T, Bruce R, Bell M, Martin M. Learning subjective language. Learning subjective language Computational linguistics. 2004; 30(3):277–308. Crossref.
8. Hu M, Liu B. Mining and summarizing customer reviews. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM New York, NY, USA; 2004. p. 168–77. Crossref.
9. Liu B, Hu M, Cheng J. Opinion observer: analyzing and comparing opinions on the web. Proceedings of the 14th International Conference on World Wide Web. Chiba, Japan; 2005. p. 342–51. Crossref.
10. Cambria E, Hussain A. Sentic computing: Techniques, Tools, and Applications, Springer: Dordrecht, Netherlands; 2012. Crossref.
11. Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP). Vancouver: Omnipress Inc; 2005. p. 347–54. Crossref.
12. Bravo-Marquez F, Mendoza M, Poblete B. Meta-level sentiment models for big social data analysis. Knowledge-Based System. 2014; 69(1):86–99. Crossref.
13. Vinodhini G, Chandrasekaran RM. Opinion mining using principal component analysis based ensemble model for e-commerce application. CSI transactions on ICT. 2014 Nov; 2(3):169–79.
14. Bai X. Predicting consumer sentiments from online text. Decision Support System. 2011; 50(4):732–42. Crossref.
15. Montoyo A, Martínez-Barco P, Balahur A. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. Decision Support System. 2012 Nov; 53(4):675–9. Crossref.
16. Basari ASH, Hussin B, Pramudya IG, Zeniarja J. Opinio n Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. Procedia Engineering. 2013; 53:453–62. Crossref.
17. Chen L, Liu C, Chiu H. A neural network based approach for sentiment classification in the blogosphere. Journal of Informetrics. 2011; 5(2):313–22. Crossref.
18. Ravi K, Ravi V. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based System. 2015; 8:14–46. Crossref.
19. Instituto Cervantes. El Espa-ol: Una lengua viva. Informe; 2014. p. 1–63.
20. Nirmal VJ, Amalarethinam DG. Emoticon based sentiment analysis using parallel analytics on hadoop. Indian Journal of Science and Technology. 2016; 9(33):2–9. Crossref.
21. Wang H, Can D, Kazemzadeh A, Bar F, Narayanan S. A system for real-time twitter sentiment analysis of 2012 U.S. Presidential election cycle. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics ACL System Demonstrations. Association for Computational Linguistics, Jeju, Republic of Korea; 2012. p. 115–20.

22. Pla F, Hurtado L-F. Political tendency identification in twitter using sentiment analysis techniques. Proceedings of COLING 25th International Conference on Computational Linguistics: Technical Papers. Dublin, Ireland; 2014. p. 183–92.
23. Purita G. Análisis de las Tendencias de uso y participación en las redes sociales a nivel Mundial en Espa-a. Obs Social; 2015. PMid:25405801
24. Haddi E, Liu X, Shi Y. The role of text pre-processing in sentiment analysis. Procedia Computer Science. 2013; 17:26–32. Crossref.
25. Team RC. R: A language and environment for Statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2014.
26. Strapparava C, Mihalcea R. SemEval-2007 Task 14: Affective text. Proceedings of the 4th International Workshop on Semantic Evaluations; 2007 Jun. p. 70–4. Crossref.
27. Mohammad S. From once upon a time to happily ever after: Tracking Emotions in Novels and Fairy Tales; 2011 Jun. p. 105–14.
28. Freelon DG. ReCal: Intercoder Reliability calculation as a web service. International Journal of Internet Science. 2010; 5(1):20–33.
29. Plutchik R, Kellerman H. A general psychoevolutionary theory of emotion. Emotion. theory, research, and experience: Theories of emotion. New York: Academic. 1980; 1:3–33. Crossref.