Sentiment Analysis in Brazilian Portuguese Tweets in the Domain of Calamity: Application of the Summarization Method and Semantic Similarity in Polarized Terms

Ariana Moura da Silva[1], Rodrigo da Matta Bastos[1] and Ricardo Luis de Azevedo da Rocha[2]
[1]*Computer Engineering Department, University of Sao Paulo, Av. Prof. Luciano Gualberto, tv. 3, 158, São Paulo, Brazil*
[2]*Languages and Adaptive Techniques Laboratory, University of Sao Paulo, São Paulo, Brazil*

Keywords: Sentiment Analysis, Natural Language Processing, Polarization, Summarization, Latent Semantic Analysis.

Abstract: This research integrates an interdisciplinary project which mobilizes the areas of Computer Engineering, Linguistics and Communication to perform the processing of texts in a natural language extracted from microblogging service Twitter as well as to conduct an analysis and classification of the sentiments mined. Many proposals have been formulated using the polarization method; however, most projects do not encompass an automatic classification by semantic proximity. This research aims to evaluate the reaction of individuals shared in the social network, not only to classify them as positive or negative, but also to ascertain the semantic similarity of these messages in the same domain. Based on the set of tweets in Portuguese extracted from a corpus of calamity, we apply three methods: a) the lexical classifier, called Summarization Method; b) the semantic classifier, called LSA - Latent Semantic Analysis; c) the ASSTPS classifier - Analysis of Semantic similarity in Polarized and Summarized terms. The results are applied to a set of 811 tweets of the calamity domain and point out which method obtained the best hit rate and semantic approximation. In this sense, the classification of sentiments by semantic proximity can help greatly, performing the sorting of content of relevant messages, discarding unnecessary information, linking messages with the same theme in common, and even generating Metrics for classifying emotions.

1 INTRODUCTION

The insertion of the World Wide Web proposed by Berners-Lee in 1990 provided the sharing of information through the first web browser (Berners-Lee, 2012). This feature provided the sharing of documents published on the Internet. Later, there was the availability of these contents in what is called the WEB 1.0, considered static. Then Web 2.0 came, and it was considered dynamic, by allowing end user interaction with the structure and content of the page, inserting comments, sending photos and sharing thousands of files. With the evolution of Web 1.0 to Web 2.0, the user went on to perform lower download rates and increased the upload rates. With the increase in the collaboration of users in Web 2.0, the need to search for such content placed on the internet rose. Web 3.0 was then established to facilitate information mining and to create a more effective language exchange between man and machine, also called the Semantic Web (Semantic Web, 2018).

The information and opinions entered by the users generated interest in the industry, by companies and financial institutions, among others, in interpreting the opinion about a product, a subject, a theme. However, reducing the large mass of data available on the WEB is a great problem, which has stirred a new area of computer science studies called sentiment analysis, which has gained prominence as from 2000 (Liu, 2012). To conduct research on sentiments analysis, it is necessary to use Natural Language Processing - NLP techniques.

Human language is not merely the manifestation of a physical action by the human being. Words are like symbols, their meanings indicate an idea or a thing. Language symbols can be encoded in voice, gesture, writing and others. NLP has different levels, from speech processing to semantic interpretation and discourse processing. NLP aims to design, to build algorithms capable of helping the machine

Silva, A., Bastos, R. and Rocha, R.
Sentiment Analysis in Brazilian Portuguese Tweets in the Domain of Calamity: Application of the Summarization Method and Semantic Similarity in Polarized Terms.
DOI: 10.5220/0006947802250231
In *Proceedings of the 10th International Joint Conference on Computational Intelligence (IJCCI 2018)*, pages 225-231
ISBN: 978-989-758-327-8

understand natural human language (Chaubard et al., 2018).

An emotion is not simply a state of sentiment (Plutchik, 2001). Sentiment analysis at the word level verifies the polarity of this word specifically. At the sentence level, not only the polarity of words is considered, but also the relationship between these words and their grammatical usage. At the document level, it considers the full context of the document, leading to a more complex analysis of how the phrases interact with one another (Liu, 2012).

According to data obtained from the Internet World Stats (Internet World Stats, 2018), Brazil is the fourth country in the world ranking of the number of users accessing the internet until December 2017. In the third position comes the USA, in the second position, India, and in the first position, China. It also tells us that the Portuguese language is in the fifth position of the most widely used language on the Internet, following Arabic, Spanish, Chinese and English in the first position (Internet World Stats, 2018). Portuguese is the fifth most used language on Twitter (The Statistics Portal, 2018), which currently has 330 million of active users in the world (The Statistics Portal; Twitter NIC, 2018).

With the exponential growth of social media, users are no longer based on the opinion of people close to them. With a Web available and easy to access, users search for and find opinions on the most diverse subjects from people around the world. The same is true when you want to find a product and think about issues from any part of the world. Thinking about structuring data, the present work is part of a slice of a larger research project, which performed the automatic extraction of tweets in Brazilian Portuguese from Twitter, used NLP techniques to the notation of the corpus, assembled a database of the domain of calamity using techniques for the construction of ontology. With the annotated corpus, two methods were applied: summarization methods of the terms (Freitas, 2015) and the LSA-latent semantic analysis (Catae, 2012).

A method is proposed from the result generated in the two previous methods. We called it semantic similarity in polarized and summarized terms - ASSTPS. The details of this method are shown in section 2. Section 3 shows works similar to the one proposed here, sentiment analysis of tweets in Portuguese. Section 4 shows a set of data in which the methods have been validated, this set may be of any other specific domain, in this work data from a corpus of calamity was used. Section 5 discusses the results found with the proposed new method and section 6 explains the conclusions and future works proposed.

1.1 Contribution

Many works have performed methods of summarization, polarization and LSA separately; it is important to emphasize that the method employed in this work is unprecedented. Here there will be a proposal of a triangular matrix which indicates the largest semantic similarity of tweets of a given domain versus the description of the synset (set of terms with the same meaning, as defined below) more semantically similar to the tweet. In addition, in this work we perform the complete translation of Sentiwordnet (Esuli et al., 2006), including the description of the synsets. The details of this method will be shown in section 2.

1.2 Goals

The objective of this work is to find an effective method to polarize terms, not in an isolated way, but rather, in such a way that the crossover of tweets with synsets and its descriptions of the Sentiwordnet are as similar as possible so that the classification is performed automatically without the manual intervention of the man in the pre-classification of similar messages. This research aims to evaluate the reaction of individuals shared in the social network, and not only to classify them as positive or negative, but also to ascertain the semantic similarity of these messages from the same domain.

2 MATERIALS AND METHODS

Messages were extracted automatically from Twitter and stored in MySQL database, composing the capture phase. Traditional NLP techniques were used for the cleanliness and notation of the corpus, following the main steps as follows: removal of stop words (words which are not aggregators of sentiment, in this work), tokenization, stemming, and tagging; composing the preprocessing phase. For a third phase, we named the classification of sentiments between positive, negative and neutral, polarity of words, using the database of Sentiwordnet (Esuli et al., 2006). This lexicon feature has 117.374 entries coming from automatic annotations from all synsets of WordNet 3.0. Synset is a set of terms with the same meaning.

Sentiwordnet is a tool which assigns a note to the degree of positivity, negativity and objectivity of

words, available only in English. Compared to the lexical resources available in Portuguese, Sentiwordnet contains a greater number of words, and is a very used resource for the classification of sentiments. Because of this we made the decision to translate the whole database of Sentiwordnet to Portuguese automatically, including the synsets and the comments which composed it.

2.1 Summarization Method

The Summarization Method or Term Score Summation (Hamouda and Rohaim, 2011), performs the sum of the positive and negative items found; for example, in the sentence below, the positive words with their values are: flowered - 0,125 and enchanted - 0,375, totalling 0,5, and the negative word with its value: polluted - 0,375. In this case, it would be categorized as a positive phrase. Algorithm 1 - "term score summation", from Sousa (2016, page 26) describes the workings of the algorithm implemented in this work.

Sample: "the girl drew the polluted river, the
neutral neg.0,375 neutral
flowered garden and the enchanted sky"
pos. 0,125 neutral pos.0,375 neutral

Sousa (2016) applies the same summarization method to analyze sentiments in texts extracted from comments from the site Tripadvisor, which is related to tourism. In his work, he proposes a second method in which, after doing the sum of the terms found in the synsets of Sentiwordnet, the average of the positive values and the average of the negative values is calculated and this value is assigned to the word (Sousa, 2016). If we use only the Summarization method, or the method of summarization with application of the average, after the translation of the synset to Portuguese, we will have a problem, as shown in the Figure 1. After the translation of the synset, we obtained a result of 12 synset "feliz" for descriptions of distinct synsets in its original form in English.

Synset Terms	Gloss	Translated Synset T	Translated Meaning
glad#4 beaming#1	cheerful and bright; "a beaming smile"; "a glad May morning"	"feliz"radiante"	alegre e brilhante; "um sorriso radiante"; "uma manhã feliz de maio"
felicitous#1	exhibiting an agreeably appropriate manner or style; "a felicitous speaker"	"feliz"	exibindo uma maneira ou estilo agradavelmente apropriado; "um orador feliz"

Figure 1: Example Synset "feliz" translated.

Thinking about this problem, we will explain what the LSA method is in the subsection below, and in section 2.3 it will be shown that the proposed new method does not perform the average of the summarization. We solved the problem by the semantic similarity of the tweet with the detailed description of the Synset.

2.2 LSA Method

Catae (2012) says that the Latent Semantic Analysis "is a technique in natural language processing, which aims to simplify the task of finding words and sentences similarity. Using a vector space model for the text representation, it selects the most significant values for the space reconstruction into a smaller dimension" (Catae, 2012).

As parameters to entry of data in this algorithm, 811 tweets will be confronted in one space sample, where the same tweets are lines and columns, so that we can visualize the similarity among all tweets.

2.3 ASSTPS Method

The ASSTPS method uses the LSA algorithm in the similarity calculation between the tweet and the description of the translated synset, to select the values to use in summarization. For example, in the case of the word "love", if the description of greater similarity is "a loved one: used as terms of affection", the values of the word "love" will be pos = 0125 and neg = 0.

For each token in the tweet, its synsets are fetched. The LSA is made between its synsets and the tweets to generate the semantic space. Within that space, the similarity between the tweet being analyzed and the synsets is calculated, allowing the choice of the most similar synset. Note that a new semantic space is generated for each token.

In this way, the result has greater accuracy, since we are not doing the sum or average of the terms, including those that are not similar. In this new method we are crossing the synsets which have high similarity for their description translated into Portuguese with the tweets of the corpus of calamity.

3 RELATED WORK

In recent years, the interest in building of sentiment classifiers has increased. Sentiment Analysis using machine learning algorithms and Bayesian networks using texts extracted from Twitter in the Spanish language was proposed by Grigori et al. in 2010 (2010). In 2014, Kanavos et al., (2014) proposed a model of sentiment analysis which measures the influence of the user on the network with the other nodes of that network. In 2013, Porshev et al., (2013) proposed a model of predicting the

psychological state of users from the texts of Twitter, and classified the forecasts using the DJIA algorithm, classifying them into eight emotions.

There are three lines of Research in the analysis of sentiment: the lexical approach, the one by machine learning and the hybrid which is the junction of the previous two (Wilson et al., 2005). We are quoting the works which have characteristics in common with the proposed work. Works which also use NLP techniques in the Portuguese language to conduct sentiment analysis. In 2017, Santos proposed a methodology for identifying polarity in Texts based on Brazilian law projects and applied it in a data set which reports on projects of law against and in favor of the liberalization of abortion in Brazil (Santos, 2017).

Freitas and Vieira (2015) present a semantic polarity classifier for the classification of comments written in Portuguese taken from TripAdvisor, an online portal of opinions on travel and accommodations (Freitas, 2015). Duarte (2013) used NLP techniques to extract mentioned entities of tweets in Portuguese and performed analysis of sentiment taking into consideration the grammatical construction of the message (Duarte, 2013). Dosciatti and Ferreira (2013) used Support Vector Machines (SVM) to identify emotions in texts written in Portuguese in Brazil, the Corpus used in the experiment is composed of news extracted from an online newspaper (Dosciatti and Ferreira, 2013). Lopes Rosa (2015) ranked the intensity of sentiments in positive, negative and neutral polarities, by the means of a new dictionary of words in Portuguese and a new calculation of sentiments (Lopes Rosa, 2015).

4 CORPUS AND SELECTION OF TEXTS FOR ANNOTATION

To perform automatic extraction of texts, it is necessary to first choose the ontology using techniques of NLP. And for storing the large amount of information we use computational linguistic techniques (CL). One of the main types of storage types in CL is in the singular corpus and in the plural body. "Corpus is a collection of language portions that are selected and organized according to explicit linguistic criteria, in order to be used as a sample of the language" (Percy et al., 1996).

A slice of data has been removed from a corpus of calamity. Calamity represents an originally positive situation and becomes a negative situation. Actions can be taken, and again it is possible that the situation which was negative becomes positive (Segers et al., 2017).

Soon it is believed that sentiment analysis in this set data becomes better validations. Although the method can be applied in any set of data, for this work we chose tweets from September 20, 2017, which were part of the topTrend '#20deSetembro' or containing the word 'Mexico' in the description of the topTrend and should still contain the Word 'earthquake' from the description of the tweet. Performing such query on the corpus returned 811 distinct tweets.

5 EVALUATION AND RESULTS

Figure 2 and the examples first show the validation of the Summarization Method alone in the calamity corpus. A total of 811 tweets, 6.789 tokens were accounted. The positive value of 284.23 shown in Figure 2 equates the sum of the lemmas values found in the automatic translation of Sentiwordnet. The negative value of 347.54 shown in Figure 2 equates the sum of the lemmas values found in the automatic translation of Sentiwordnet.

375 tweets were classified for positive polarity and 428 for negative polarity. It can therefore be concluded that the issue in question "Earthquake in Mexico on 20 September 2017" obtained 52.77% of negativity, 46.24% of positivity.

Total rows	**811**	Qty **rows** with **Neutral** Qty	**8**
Total **Positive**	**284.2367**	Total **Negative**	**347.5466**
Average Positive	0.3505	**Average** Negative	0.4285
Qty **rows** with **Positive** Polarity Qty	**375**	Qty rows with **Negative** Polarity	**428**

Figure 2: Results Summarization Method.

Samples from the corpus which were sorted correctly by the Summarization Method were extracted below. The samples are in the Portuguese language.

Sample with 1,2867 positive polarity: "*Visitei o México ano passado, um país sensacional, cheio de gente feliz, tô mal com a notícia do terremoto*".
Sample with 1,7806 negative polarity: "*México é um belo País! Que tristeza o Terremoto de ontem. E agora, Porto Rico em perigo novamiente tambem nao me deixa tranquila aqui!*".

Figure 3 is a sample of the data for the validation of the LSA Method, where high rates of semantic similarity were found in tweets compared two to two with opposite polarities. At 0.53 rate of similarity for the tweet ID: 4 and 446; and the rate of 0.52

similarity to the tweets ID: 671 and 765. Figure 4 is a sample of the data for the validation of the LSA Method, where high rates of semantic similarity were found in tweets compared two to two with equal polarities. At 0.61 rate of similarity for the tweet ID: 29 and 47; and the rate of 0.72 similarity to the tweets ID: 80 and 431.

According to the results from the first two methods, discrepancy was observed, not allowing a fundamental conclusion for the event of September 20, 2017 with the natural phenomenon of an earthquake in Mexico. The results should obviously be of sadness, unhappiness and anguish, pointing to a negative polarity analysis.

ID Tweet	Description	Pos. Value	Neg. Value	Polarity	Similarity -1 until +1
29	Terremoto no México mata mais de 100 pessoas: .	0.1474	**0.1941**	**Negative**	**0.6075**
47	Gente em 19 de Setembro de 1985 também houve um terremoto no México e matou milhares de pessoas. Seria cômico se não fosse trágico.	1.0119	**1.5518**	**Negative**	
80	População participava de treinamento para lidar com abalos sísmicos no momento do terremoto	**0.3333**	0.1845	**Positive**	**0.7235**
431	Terremoto na Cidade do México acontece no momento em que pessoas faziam um treinamento para lidar com sismos	**0.5708**	0.1054	**Positive**	

Figure 3: Results - High Rate Similarity x Polarity Opposite.

ID Tweet	Description	Pos. Value	Neg. Value	Polarity	Similarity -1 until +1
4	Visitei o México ano passado, um país sensacional, cheio de gente feliz, tô mal com a notícia do terremoto	**1.2867**	1.1973	**Positive**	**0.5327**
446	Estivemos na Cidade do México há 2 anos, em set/2015. É triste ver as notícias do terremoto que atingiu a região on	0.3261	**0.8955**	**Negative**	
671	México é um belo País! Que tristeza o Terremoto de ontem . . . E agora, Porto Rico em perigo novamente tambem nao me deixa tranquila aqui!	1.0554	**1.7807**	**Negative**	**0.5242**
765	E enviemos nossas orações aos povos do México, da N. Zelândia e de Porto Rico. Os primeiros sofreram com terremotos; o último, com furacão.	**0.4501**	0.3494	**Positive**	

Figure 4: Results - High Rate Similarity x Polarity Equal.

We sought some common synsets in the field of calamity, for example: Earthquake, volcano and cyclone. And we realized that most of them had zero values in the Sentiwordnet database. Figure 5 shows examples of common synsets in the calamity corpus with zero values. And one of the synsets "earthquake" is of positive value of 0.125. This caused the new applied method to show high values for positive polarity, since the term "earthquake" appears in all the tweets analyzed in this work. Two analyses were performed for validating the ASSTPS method. The first, as shown in Figure 6, kept the original value found for the synset "earthquake" in the Sentiwordnet database; the second analysis, as shown in Figure 7, changed to 0.25 the negative value for all the "earthquake" synsets found.

For validating the ASSTPS method, the synset values found in Sentiwordnet remained; the result of the method here proposed categorized in the base 449 tweets with positive polarity, representing 55.36% of the base.

The earthquake is a natural disaster event, and its causalities stir actions of dissatisfaction in the population. Thus, as a complementary exercise, we adjusted the negative value of the Synsets "earthquake" to 0.25. We re-processed the algorithm and Figure 7 shows the values obtained from 647 tweets classified by proximity semantics in the description of the synset, representing 79.77%.

Synset Terms	Gloss	Pos.	Neg.
temblor#1 seism#1 quake#1 earthquake#1	shaking and vibration at the surface of the earth resulting from underground movement along a fa	0	0
tremor#1 quake#2	shake with seismic vibrations; "The earth was quaking"	0.125	0
quiver#1 quake#1 palpitate#2	shake with fast, tremulous movements; "His nostrils palpitated"	0	0
volcano#1 vent#3	a fissure in the earth's crust (or in the surface of some other planet) through which molten lava an	0	0
volcano#2	a mountain formed by volcanic material	0	0
vulcan#1	(Roman mythology) god of fire and metal working; counterpart of Greek Hephaestus	0	0
cyclone#2	a violent rotating windstorm	0	0.125
anticyclone#1	(meteorology) winds spiraling outward from a high pressure center; circling clockwise in the north	0	0
cyclone#1	(meteorology) rapid inward circulation of air masses about a low pressure center; circling counter	0	0

Figure 5: Example Calamity Synsets.

Total rows	**811**	Qty **rows** with **Neutral** Qty	**118**
Total **Positive**	**236.0000**	Total **Negative**	**234.2500**
Average Positive	0.2910	**Average** Negative	0.2888
Qty **rows** with **Positive** Polarity Qty	**449**	Qty rows with **Negative** Polarity	**244**

Figure 6: Results – ASSTPS Method.

Total rows	**811**	Qty **rows** with **Neutral** Qty	**79**
Total **Positive**	**139.0000**	Total **Negative**	**426.7500**
Average Positive	0.1714	**Average** Negative	0.5262
Qty **rows** with **Positive** Polarity Qty	**85**	Qty rows with **Negative** Polarity	**647**

Figure 7: Results – ASSTPS Method with synset “earthquake” negative value 0.25.

6 CONCLUSIONS AND FUTURE WORK

One of the main contributions of this work is the research into sentiment analysis performed in a Brazilian Portuguese calamity corpus. In this work, we present semantic classifications of polarity for analyzing texts. The classifiers used Sentiwordnet to assign values to words. The entire collection of Sentiwordnet was automatically translated into Portuguese; even with translation errors, the classifiers that used the translated Sentiwordnet achieved good performance.

The Summarization method is a lexical classifier that has some mismatches. The LSA method performs the semantic proximity of the tweets among other tweets and does not perform any polarity check. The proposed new method performs the semantic similarity of the tweets with the synsets of Sentiwordnet by semantic proximity and performs the polarity classification. The polarization of sentiments by semantic proximity associated with the synsets of Sentiwordnet and their comments can be of great help in tasks, such as sorting content of relevant messages, discarding unnecessary

information, linking messages with the same theme in common, and even generating metrics for classifying emotions.

As future works, we suggest improvements to the translation process, performing bi-gram and tri-gram words as proposed in the work by Lopes Rosa (2015). Use the new method proposed to classify emotions at a second level, such as anger, fear, love and hate; above all, to use a base similar to Sentiwordnet with values of positivity, negativity and more accurate objectivity for the dominance of calamity.

ACKNOWLEDGEMENTS

The authors are grateful to the CNPQ process 141077/2015-8 for the support received.

REFERENCES

Berners-Lee, Tim. The WorldWideWeb browser. World Wide Web Consortium. Homepage, https://www.w3.org/People/Berners-Lee/WorldWide Web.html, last accessed 2018/03/ 01.

Catae, F. S. Classificação Automática de Texto por Meio de Similaridade de Palavras: um algoritmo mais eficiente. *Master's Dissertation*, Universidade de São Paulo (2012).

Chaubard, F., Fang, M., Genthial, G. et al. Natural language processing with deep learning. Lecture Notes: Part I - Winter - 2017. Homepage, https://tensorflowkorea.files.word press.com/2017/03/ cs224n-2017winter-notes-all.pdf, last accessed 2018/04/15.

Dosciatti, M. M.; Ferreira, E. C. L. P. C. Identificando emoções em textos em português do brasil usando máquina de vetores de suporte em solução multiclasse. *ENIAC-Encontro Nacional de Inteligência Artificial e Computacional*. Fortaleza, Brasil (2013).

Duarte, E. S. Sentiment analysis on twitter for the portuguese language. PhD Thesis, Faculdade de Ciências e Tecnologia (2013).

Esuli, A., Sebastiani, F. Sentiwordnet: A publicly available lexical resource for opinion mining. In: In *Proceedings of the 5th Conference on Language Resources and Evaluation - LREC'06*, 417-422 (2006).

Freitas, L. D; Vieira, R. Exploring resources for sentiment analysis in portuguese language. In: *IEEE. 2015 Brazilian Conference on Intelligent Systems (BRACIS)* (2015).

Freitas, L. A. de. Feature-level sentiment analysis applied to brazilian Portuguese reviews. PhD Thesis, Pontifícia Universidade Católica do Rio Grande do Sul (2015).

Grigori, S. et al. Empirical study of machine learning based approach for opinion mining in tweets. In *Proceedings of the 11th Mexican international conference on Advances in Artificial Intelligence* (2012).

Hamouda, A., Rohaim, M. Reviews classification using sentiwordnet lexicon. In: *World Congress on Computer Science and Information Technology* (2011).

Internet World Stats. Top 10 Languages Used in the Web - December 31, 2017. Homepage, https://www.internetworldstats.com/stats7.htm, last accessed 2018/04/01.

Internet World Stats. Top 20 Countries With the Highest Number of Internet Users - December 31, 2017. Homepage, https://www.internetworldstats.com/ top20.htm, last accessed 2018/02/01.

Kanavos, A. et al. Conversation Emotional Modeling in Social Networks, Tools with Artificial Intelligence (ICTAI), *2014 IEEE 26th International Conference* (2014).

Liu, B. Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies. California: Morgan & Claypool Publishers (2012).

Lopes Rosa, R. Análise de sentimentos e afetividade de textos extraídos das redes sociais. Tese de Doutorado – Escola Politécnica da Universidade de São Paulo (2015).

Percy, C. E. et al. Synchronic Corpus Linguistics – Papers from the *sixteenth International. Conference on English Language and Research on Computerized Corpora (ICAME 16)*, Amsterdam/Atlanta (1996).

Plutchik, R. The nature of emotions. v. 89, n. 4, p. 344–350 (2001).

Porshnev, A.; Redkin, I.; Shevchenko, A., "Machine Learning in Prediction of Stock Market Indicators based on Historical Data and Data from Twitter Sentiment Analysis," *Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference* (2013).

Sousa, R. C. C. Identificando Sentimentos de Textos em Português com o Sentiwordnet traduzido. Monografia – UFC, curso de Ciência da Computação (2016).

Santos, D. L. B. Metodologia de Identificação de Polaridade em Textos com Base em Projetos de Lei Brasileiros. Dissertação de mestrado – UFRJ, Programa de Engenharia Civil (2017).

Segers, R., Caselli, T., Vossen, P. The Circumstantial Event Ontology (CEO). *Proceedings of the Events and Stories in the News Workshop*. Association for Computational Linguistics, Vancouver, Canada (2017).

Semantic Web - W3C. Semantic Web. Homepage, https://www.w3.org/standards/semantic web/, last accessed 2018/03/01.

The Statistics Portal. Most-used languages on Twitter as of September 2013. Homepage, https://www.statista.com/statistics/267129/most-used-languages-on-twitter/, last accessed 2018/04/08.

The Statistics Portal. Most popular social networks worldwide as of April 2018, ranked by number of active users (in millions). Homepage, https://www.statista.com/statistics /272014/global-social-networks-ranked-by-number-of-users/, last accessed 2018/04/08.

Twitter NIC. A Multilingual Social City. Homepage, http://ny.spatial.ly/, last accessed 2018/04/25.

Wilson, T.; Wiebe, J.; Hoffmann, P. Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: *Association for Computational Linguistics* (2005).