Journal of Universal Computer Science, vol. 22, no. 5 (2016), 671-690
submitted: 14/9/15, accepted: 29/4/16, appeared: 1/5/16 © J.UCS

Web Service SWePT: A Hybrid Opinion Mining Approach

Yolanda Raquel Baca-Gomez
(Center for Research and Innovation in Information and Communications Technologies, Mexico City, Mexico
yolanda.baca@infotec.mx)

Alicia Martinez
(National Center for Research and Technology Development
Cuernavaca, Morelos, Mexico
amartinez@cenidet.edu.mx)

Paolo Rosso
(PRHLT, Technical University of Valencia, Valencia, Spain
prosso@dsic.upv.es)

Hugo Estrada
(Center for Research and Innovation in Information and Communications Technologies, Mexico City, Mexico
hugo.estrada@infotec.mx)

Delia Irazu Hernandez Farias
(PRHLT, Technical University of Valencia, Valencia, Spain
dhernandez1@dsic.upv.es)

Abstract: The increasing use of social networks and online sites where people can express their opinions has created a growing interest in Opinion Mining. One of the main tasks of Opinion Mining is to determine whether an opinion is positive or negative. Therefore, the role of the feelings expressed on the web has become crucial, mainly due to the concern of businesses and government to automatically identify the semantic orientation of the views of customers or citizens. This is also a concern, in the area of health to identify psychological disorders. This research focuses on the development of a web application called SWePT (*Web Service for Polarity detection in Spanish Texts*), which implements the Sequential Minimal Optimization (SMO) algorithm, extracting its features from an affective lexicon in Mexican Spanish. For this purpose, a corpus and an affective lexicon in Mexican Spanish were created. The experiments using three *(positive, neutral, negative)* and five categories *(very positive, positive, neutral, negative, and very negative)* allow us to demonstrate the effectiveness of the presented method. SWePT has also been implemented in the Emotion-bracelet interface, which shows the opinion of a user graphically.

Keywords: Opinion Mining, affective lexicon, hybrid approach, web service
Categories: C.3, J.4, H.3.1, H.3.5, H.5.2, I.2.6, I.2.7

1 Introduction

Opinion Mining is a broad area of Natural Language Processing and text mining, which is defined as the computational study of opinions expressed in texts regarding entities such as products, services, organizations, individuals, issues, topics, and their attributes [Liu, 2012; Ortigosa-Hernandez, 2012]. The major task of Opinion Mining is the classification of the opinion's polarity or semantic orientation into different categories, which are usually designated as positive, neutral, and negative. This classification consists of determining whether the opinion is positive, negative, or neutral with respect to the entity to which it is referring (a person, a product, a movie, etc.) [Balahur, 2012; Martin-Wanton, 2010].

Opinions are central to almost all human activities because they are key influencers of our behaviors; whenever we need to make a decision, we want to know other's opinions [Liu, 2012]. Hence, this area of research is becoming more and more important mainly due to the growth of social media where users continually share opinions, sentiments, evaluations, attitudes, and emotions [Liu, 2012; Martin-Valdivia, 2013]. However, the number of opinions shared on the web has increased exponentially and it is becoming an impossible task to read all of these opinions. Therefore, it is necessary to develop new methods for the automatic processing of opinions [Pang, 2002; Rushdi, 2011; Poria, 2014].

Most studies on polarity classification only deal with English documents, mainly due to the lack of resources in other languages [Martin-Valdivia, 2013; Fernandez-Anta, 2013]. Despite the fact that Spanish is among the top ten languages that are most used on the Internet according to the Internet world state rank[1], there are few resources for managing sentiments or opinions in this language. Consequently, there is an increasing need for the study of Opinion Mining in languages other than English [Martin-Valdivia, 2013].

The main purpose of this work is the development a web service for polarity detection in Mexican Spanish. Through the implementation of a method that is based on a hybrid approach. This method combines the Sequential Minimal Optimization (SMO) machine learning algorithm with the use of features obtained by an affective lexicon in Mexican Spanish and a corpus. The tests for the method proposed were carried out in three ways: 1) by comparing the results taking into account different lexical features using a general corpus; 2) by comparing the results with an already existing affective lexicon using a general corpus; and 3) by comparing the results with the Wolfram Mathematica system [Wolfram, 2016] using a domain-specific corpus.

The paper is organized as follows: Section 2 presents some related work; Section 3 describes the SWePT method, including a general view of our proposal, the resources generated [Baca-Gomez, 2014], and the method architecture; Section 4 shows the experiments and results obtained; Section 5 describes the case study; and Section 6 presents the conclusions and the future work that can be derived from this work.

---

[1] Source: Internet World Stats - http://www.internetworldstats.com/stats7.htm. Estimated Internet users are 2,802,478,934 on December 31, 2013

2 Related work

In general, the approaches described in this work can be classified into three categories: (i) machine learning methods, (ii) lexicon-based methods, and (iii) hybrid methods.

2.1 Machine learning methods

The machine learning methods, such as Naïve Bayes, Support Vector Machine, or Artificial Neural Networks have been used for affect classification of texts by feeding a machine learning algorithm with a large training corpus of affectively annotated texts. Generally, these methods only work with acceptable accuracy with sufficiently large text inputs [Poria, 2014]. They may suffer from overfitting and are highly dependent on the quality, size, and domain of the training data [Martin-Wanton, 2010].

The majority of machine learning text classification approaches employ Support Vector Machine classifiers, which are trained on a specific data set. These use features such as unigrams or bigrams (with or without part of speech tagging), although the most successful features seem to be basic unigrams [Taboada, 2011].

One of the most representative works in this category is the one described in [Pang, 2002]. The aim of this work was to examine the feasibility of treating sentiment classification as a case of topic-based categorization where the topics are positive, negative or whether other aspects must be taken into account (e.g. bigrams, part of speech tagging, word position, and feature frequency vs presence). The three machine learning methods used in that work were Naïve Bayes, Maximum Entropy, and Support Vector Machine. The best results were obtained with Support Vector Machine. In the works described in [Martinez-Camara, 2011, Martinez-Camara, 2011b] some experiments were made using a Support Vector Machine and Naïve Bayes. The aim of these two works was to evaluate algorithms that focus on texts in Spanish. The best results for both works were obtained using the Support Vector Machine.

Another approach is Deep Learning, in which the studies focus on learning the low-dimensional, dense, and real-valued vector as text features for Sentiment Analysis without any feature engineering [Tang, 2015]. In the work described in [Bespalov, 2011] they built an embedding mechanism of n-grams to low-dimensional latent semantic space, where a classification function can be defined. They used a Deep Neural Network to build a unified discriminative framework that allows for estimating the parameters of the latent space as well as the classification function with bias for the target classification task at hand. Also, in the work presented in [Glorot, 2011] they proposed a Deep Learning approach, which learns to extract a meaningful representation for reviews in an unsupervised fashion. They use Stacked Denoising Autoencoder, which is a kind of Neural Network that is optimized by reconstructing the input itself. Denoising Autoencoding randomly masks the values of inputs and tries to reconstruct the noisy inputs.

More recently, some studies suggest that ensemble learning methods may have potential applicability in sentiment classification [Xia, 2011; Wang, 2014]. Wang [Wang, 2014] presents a comparative assessment of the performance of three popular ensemble methods, Bagging, Boosting, and Random Subspace. These methods are based on five learners for sentiment classification, Naïve Bayes, Maximum Entropy, Decision Tree, K Nearest Neighbor, and the Support Vector Machine. The results obtained in the works by Wang illustrate that ensemble methods can be used as a viable method for sentiment classification.

2.2 Lexicon-based methods

Lexicon-based methods use dictionaries of sentiment words and phrases with their associated orientations and strength. The lexicon usually incorporates intensification and negation to compute a sentiment score for each text. It has been shown that the lexicon-based methods perform quite well in a large number of domains [Liu, 2012]. Also, part of speech information is commonly indicated in lexicons, partly to overcome word sense disambiguation, which therefore helps to achieve better sentiment classification performance [Gezici, 2013]. However, this method also has its own limitations. Besides sentiment words and phrases, there are many other types of expressions that involve sentiments and most of them are difficult to handle [Liu, 2012].

One of the most important works in this category is the method presented in [Turney, 2002], where a learning algorithm was implemented for classifying reviews as *recommended* (thumbs up) or *not recommended* (thumbs down). The PMI- IR algorithm (Pointwise Mutual Information – Information Retrieval) was employed to estimate the semantic orientation of a sentence. The semantic orientation of a given word was calculated by comparing its similarity to a positive reference word with its similarity to a negative reference word. Some years later, this work was replicated using Spanish texts [Cruz, 2008].

A system that was developed by the Serendio team is described in [Palanisamy, 2013]. The system is constructed on a lexicon-based approach for discovering sentiments. The lexicon is built from the Serendio taxonomy, which consists of positive/negative, affirmation/negation, stop words and phrases/sentences. The system has a pre-processing step, which includes stemming, emoticon detection, normalization, exaggerated word shortening, and hashtag detection. After the pre-processing, the system classifies tweets as positive or negative based on the contextual sentiment orientation of the words. The sentiment calculation is the aggregation of the sum of the sentiment-bearing entities of the tweet. This aggregation is based on a set of heuristics that is built on the sentiment orientation of the words.

In the work presented in [Benamara, 2007], the proposed method is an Adverb-Adjective Combinations-based Sentiment Analysis technique that uses a linguistic analysis of adverbs of degree. In a recent work by [Taboada, 2011], a system called Semantic Orientation CALculator was developed. This system extracts sentiment-bearing words, including adjectives, verbs, nouns, and adverbs and uses them to calculate semantic orientation, taking into account valence shifters (intensifiers, downtoners, negation, and irrealis markers).

2.3 Hybrid methods

Hybrid methods assemble closely related methods in order to obtain better precision than conventional approaches. Thus, a hybrid classifier might take advantage of multiple Sentiment Analysis approaches [Martin-Valdivia, 2013; Balage-Filho, 2014].

For example, a hybrid approach has been used to combine lexical analysis and machine learning to cope with ambiguity and integrate the context of sentiment terms [Martin-Valdivia, 2013; Poria, 2014].

A system for Sentiment Analysis that integrates several techniques is presented in [Saralegi, 2012]. This system adopts an approach that includes linguistic knowledge-based processing for preparing features. The processing includes the treatment of the following based on the syntactic nesting level: errors, lemmatization, part of speech tagging, tagging of polarity words, treatment of emoticons, negation, and weighting of polarity words. The detection of polarity words is done according to a polarity lexicon.

In the work presented by [Martin-Valdivia, 2013], a meta-classifier that combines supervised and unsupervised learning for polarity classification is proposed. This system uses a Spanish corpus of movie reviews along with its parallel corpus which is translated in English. First, two individual models were generated applying machine learning algorithms. Second, SentiWordNet [Baccianella, 2010] was integrated generating an unsupervised model. Then the models and SentiWordNet were combined using a meta-classifier that allows combination algorithms to be applied.

Marchand [Marchand, 2013] proposes a mixed approach for Sentiment Analysis in Twitter. First, tweets are filtered based on the occurrences of words from a sentiment lexicon, and then supervised learning methods are applied (e.g., bag of words, Support Vector Machine or Tree Kernel). To maximize efficiency, normalization steps are performed, (e.g., lemmatization and syntactic parsing).

Vilares [Vilares, 2013] presents a system that classifies the polarity of Spanish tweets. A hybrid approach is adopted, which combines machine learning and linguistic knowledge that is acquired by means of Natural Language Processing. A part of speech tagging syntactic, dependencies and semantic knowledge are used, as features of a supervised classifier.

In the works presented in [Balage-Filho, 2014], a system is presented which adopts a hybrid classification process. This system uses three classification approaches: rule-based, lexicon-based, and machine learning. The system is divided into four main components: normalization, rule-based classifier, lexicon-based classifier, and machine learning classifier. These components are connected in a pipeline architecture that extracts the best characteristics from each component.

In [San Vicente, 2014], a system for Sentiment Analysis is presented, which implements a Support Vector Machine algorithm. The algorithm combines the information extracted from polarity lexicons that have linguistic features.

3 The SWePT method

The SWePT method is a hybrid approach for polarity detection of texts in Mexican Spanish. This method combines the SMO algorithm with feature extraction from an affective lexicon. Thus, the method requires a corpus and feature extraction from text. The corpus was developed using comments from Facebook and Twitter. Therefore, an automatic extraction software system was implemented, and an affective lexicon in Mexican Spanish was created.

The SWePT method is divided into three main modules: pre-processing, feature extraction, and automatic classification. The two relevant resources created are: the affective lexicon in Mexican Spanish and the corpus. Figure 1 shows the execution flow of the modules and the resources generated. Each one of the modules and resources that are relevant for the proposed method is presented below in detail.

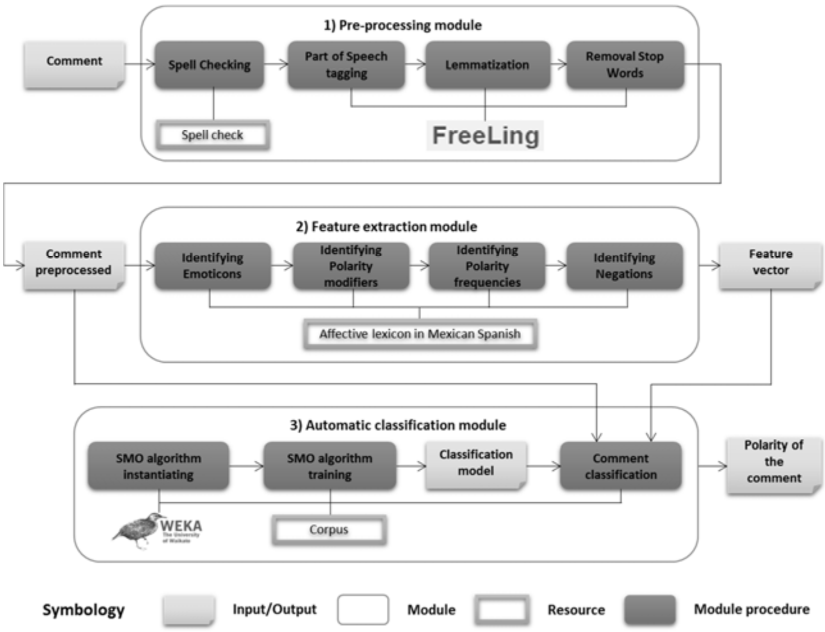

Figure 1: The modules of the SWePT method.

3.1 Pre-processing module

The input of this module is a comment. The output of this module is a preprocessing comment and is free of spelling errors and stop words (lemmatized, grammatically labeled). This module is composed of four sub-modules, the last three are implemented using *FreeLing* [Padro, 2012].

Spell Checking. In this sub-module, there is a spell checker implementing the following:

Abbreviations and chat terms are modified; for example, abcs → *a veces* (in English *sometimes*), a100do → *haciendo* (in English *doing*).

Laughs are standardized; for example, jajajajajaaaja → jaja, hahaha → jaja, jeejeje → jeje.

Repeated letters and signs are eliminated; for example, *te quieroooo → te quiero* (in English I love you), *te odio!!!! → te odio* (in English I hate you).

Web links are eliminated.

Part of speech tagging. In this sub-module, the grammatical category for each word in the text is identified. For example, the word *run* is labeled as a verb, the word *the* is labeled as an article, and the word *that* is labeled as a pronoun.

Lemmatization. In this sub-module, the lemma of a given word is determined by grouping together the different inflected forms of a word. Therefore, the objective is to obtain the root of group of words. For example, the words *ran* and *running* are converted in the word *run*.

Removal stop words. In this sub-module, the words without a relevant meaning for the analyzed text are removed. In this case, words belonging to the following categorization are eliminated: determiners, articles, pronouns, conjunctions, numbers and numerals.

3.2 Feature extraction module

The goal of this module is to extract text features that provide information that is relevant to the detection of the polarity and intensity. The module takes into account factors such as denials, polarity modifiers, and polarity expressions. As a result, the process generates a feature vector that is based on the affective lexicon in Mexican Spanish. The feature vector is generated taking into account the following sub-modules:

Identifying and grouping emoticons. In this sub-module, the emoticons are classified as positive or negative; for example, **:3** (this emoticon looks like a cat-face) → positive; **u.u** (this emoticon expressing sadness or disappointment)→ negative.

Identifying polarity modifiers. In this sub-module, the words that directly modify the polarity of a given word are identified; for example, *very happy.*

Identifying polarity frequencies. In this sub-module, the frequencies of *very positive, positive, very negative* or *negative* words are identified.

Identifying negations. In this sub-module, the words that negate the polarity of a given word are identified; for example, *I am not happy.*

3.3 Automatic classification module

The goal of this module is to identify the polarity of the comments. The input of this module is a comment pre-processed and the feature vector. This module was implemented using the *Weka* library [Hall, 2009] and the SMO algorithm. The output of this module is the polarity of the comment. Three main modules compose the automatic classification module: SMO algorithm instantiating, the SMO algorithm training and finally, the comment classification.

An example of the application of these three modules is presented below, where the following comment is used as an example: *My best friend doesn't lie to me; he protects me and keeps his promises <3 I loveeeeeeee him :3.* In this example, the phrase <3 represents a heart and the phrase *:3* represents a cat-face.

Applying the pre-processing sub-module. As a result of the pre-processing stage, the text has been divided as follows: *best friend, does not lie, protects, keeps promise, positive emoticon, love, positive emoticon.* It is important to point out that the repeated letters were eliminated. However, *love* is a positive word. Because of the repeated letters, this word is taken as very positive.

Applying the feature extraction sub-module. As a result of the feature extraction module, a feature vector was defined with the following variables: 1) the number of positive words, 2) the number of very positive words, 3) the number of negative words, 4) the number of very negative words and 5) the pre-processed comment. The words of the source comment were categorized as follow: The word *friend* is positive, and the word *best* is taken as an intensifier. Therefore, *best friend* is very positive.

The word *lie* is negative, but there is a negation. Hence, this word is taken as positive. The word *protects* is positive. The word *keeps* is positive. *Love* is a positive word. However, because of the repeated letters, it is taken as very positive.

As output we have the following feature vector: two very positive words (*best friend* and *loveeeeeee*). Three positive words (*does not lie, protect, keep)* and two positive emoticons (<3 and :3). Zero very negative words, and zero negative words.

Applying the automatic classification sub-module. Finally, the process concludes with the obtaining of the polarity, which determines that the comment "*My best friend doesn't lie to me; he protect me and keeps his promises <3 I loveeeeeeee him :3"* has very positive polarity.

3.4 Resources generated

This subsection describes the resources generated as part of the proposed method (the affective lexicon in Mexican Spanish and the corpus).

The affective lexicon in Mexican Spanish was created to obtain features that provide more information to the SMO algorithm. The followed steps for the creation of the affective lexicon are listed below:

Step 1. Manual translation from English to Spanish of words obtained from lexical resources. A set of words was translated (from English to Spanish) and labeled with their polarity and emotional category. This set words was obtained from psychological theories [Russell, 1980; Sacharin, 2012; Scherer, 2013; Scherer, 2005; Morgan, 1988] and affective lexicons [Strapparava, 2004; Stone, 1966; Wilson, 2005].

Step 2. Manual enrichment based on semantic relationships. The set of words was enriched with semantic relationships: lexical families, inclusion relationships, and synonyms.

Step 3. Manual enrichment of Mexican slang. The affective lexicon was enriched with Mexican slang and other expressions like emoticons and interjections that are commonly used on Facebook and Twitter. The meaning and the semantic orientation of the expression were also added based on the context in which the word is used.

The corpus must be labeled with categories, and these categories should be defined based on the purpose of the corpus [Sarmento, 2009; Schulz, 2010; Wiebe, 2006]. In this work, two types of corpora were established. The first one with three categories: *positive, neutral and negative.* The second one had five categories: *very positive, positive, neutral, negative,* and *very negative.*

Two experiments were performed to determine whether having information about polarity improved the results of the manual labeling. In the first experiment the corpus had five categories with two different scenarios. Three people manually labeled the comments in the corpus.

Scenario 1: In this scenario, the people did not have any knowledge about polarity. They labeled each comment of the corpus as: *very positive, positive, neutral, negative,* or *very negative.*

Scenario 2: In this scenario, the people did have previous knowledge about polarity. They labeled each comment of the corpus as: *very positive, positive, neutral, negative,* or *very negative.*

In the second experiment, a corpus with three categories was created (based on corpus created in scenario 1 and 2) by replacing the label *very positive* with the label *positive* and the label *very negative* with the label *negative*. Both corpora were validated using the Fleiss' Kappa metric [Fleiss, 2003] in order to obtain the agreement among the three people labeling the two corpora.

Table 1 shows the results of the Fleiss' Kappa evaluation. Since scenario 2 had better results, it was chosen as the labeling method for our study.

	3 categories	**5 categories**
Scenario 1	0.2035	0.1125
Scenario 2	**0.4365**	**0.3367**

Table 1: The results of the validation of the corpus in the two scenarios

3.5 Developing general and specific corpora

Once the labeling method has been defined, two corpora were made in order to determine whether a general corpus or a specific-domain corpus was better.

General corpus. Four versions of this corpus were made with three and five categories, and two sizes (with 1500 and 3100 comments). The comments were randomly obtained from social networks using our automated extraction system. Table 2 shows the composition of the general corpus with two categories and two sizes.

	Category	**Corpus with 1500 comments**		**Corpus with 3100 comments**	
		5 categories	3 categories	5 categories	3 categories
1	Very positive	274	0	470	0
2	Positive	397	671	832	1302
3	Neutral	290	290	469	469
4	Negative	282	539	746	1329
5	Very negative	257	0	583	0

Table 2: The general corpus with 1500 and 3100 comments

The four versions of this corpus were labeled by six people. The agreement among six people was evaluated using Krippendorff's alpha [Krippendorff, 2011]. Table 3 shows the results.

	3 categories	**5 categories**
1500 comments	0.635	0.411
3100 comments	0.422	0.339

Table 3: The results of the of the inter-annotator evaluating using Krippendorff's alpha

The specific-domain corpus. This corpus was made using the selected domain *prices* and it was limited to a specific product. This corpus can be used to monitor people´s behavior when there is a rise in the price of a given product. The analyzed comments were obtained using the search string *price egg* through our automated extraction system. The goal of this analysis was to compare the results when a general corpus is used in a specific domain, and when a corpus is specially created for a specific domain. Our experiments showed that when a specific behavior needs to be analyzed, then a specific-domain corpus works better. This corpus had three categories because the previous results had revealed that the agreement between people labeling is better when there are three categories. The corpus was labeled by four people and evaluated using Krippendorff's alpha. The result was 0.442.

Category	**Number of comments**
Positive	133
Neutral	209
Negative	195
Total comments	537

Table 4: The number of comments in the specific-domain corpus

3.6 The SWePT architecture

The SWePT web service is a service that is offered by a software component to a client application that consumes the web service. The client sends a call through a servlet. The servlet generates the connection and exchanges data between the client and the web service, which resides on a web server. Once the client is connected with the web service, the web service processes the comment and returns the polarity to the client application.

Figure 2 shows the SWePT web service. The step-by-step process is outlined below:

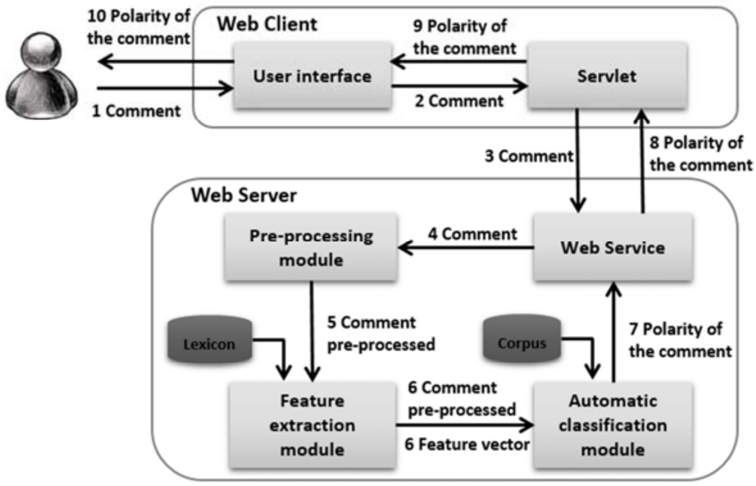

Figure 2: The SWePT Web Service architecture

1. The user posts a comment in the user interface.
2. The user interface establishes the connection between the servlet and sends the comment.
3. The servlet sends the comment to the homepage of the web service.
4. The web service sends the comment to the pre-processing module.
5. Once the comment has been pre-processed, the comment is sent to the feature extraction module.
6. The feature extraction module generates the feature vector and sends the pre-processed comment and the feature vector generated to the automatic classification module.
7. The classification module identifies the polarity for the comment and sends it back to the homepage web service.
8. The web service sends the polarity of the comment to the web client through the servlet.
9. The servlet sends the polarity of the comment to the user interface.
10. The user interface shows the polarity of the comment.

4 Experiments and results

Three evaluations were carried out to demonstrate the effectiveness of the proposed method. In the first evaluation, the SWePT method was analyzed taking into account different features in order to identify which combination of features had the best performance. In the second evaluation, a comparison was carried out using our affective lexicon in Mexican Spanish and the Spanish lexicon by Perez [Perez-Rosas, 2012]. In the third set of experiments, the SWePT method was evaluated with a specific-domain corpus. The results obtained by the SWePT method were compared with the results of the Wolfram Mathematica system using the same corpus in both cases. Wolfram Mathematica is a system to apply mathematics in different fields of knowledge. And, also incorporates a programming language that allows statistical analysis and text analysis [Wolfram, 2016].

4.1 Evaluation 1. Experiments with different features

In the set of experiments for Evaluation 1, the SWePT method was evaluated taking into account seven features in order to identify which combination of features worked best. The experiments were carried out using the ten-fold cross validation technique and the SMO algorithm implemented with the *Weka* library: without any kind of pre-processing; with pre-processing (spell checking, part of speech tagging, lemmatization, stop words removal); grouping emoticons in two groups (*positive* and *negative)*; considering polarity frequencies (number of words or expressions classified as *very positive, positive, very negative* or *negative*); polarity modifiers; negations; emotions.

The best result was obtained in Experiment 8 which considered: pre-processing, grouping emoticons, polarity frequencies, polarity modifiers, and negations. In general, the results with three categories (*positive*, *negative* and *neutral*) were the best in all cases. This is because the method often confuses the categories, *very positive* and *positive* and *very negative* and *negative*.

Nevertheless, the greater the number of categories defined, the more difficult it is to identify what characteristics belong to each category. This directly impacts the precision of the method. The results of the corpus with 1500 comments are better than the corpus with 3100 comments. This is because the first corpus has a better balance in the number of elements in each category. Also, the first corpus has a better result in the evaluation with Krippendorff's alpha. The second corpus has more positive comments that negative comments. Therefore, the method knows more the positive category than the negative category.

Table 5 shows the results with the F-Measure which can be interpreted as a weighted average for the precision and recall. The features taking into account are the following: (a) Without any kind of pre-processing; (b) Grouping emoticons in two groups: positive and negative; (c) Pre-processing (spell checking, part of speech tagging, lemmatization, stop words removal); (d) Polarity frequencies (number of words or expressions classified as: very positive, positive, very negative or negative); (e) Polarity modifiers; (f) Negations and (g) Emotions.

Experiments	**Features**	**Corpus 1500**		**Corpus 3100**	
		5 cat.	**3 cat.**	**5 cat.**	**3 cat.**
Experiment 1	(a)	51.7%	70.4%	46.9%	66.1%
Experiment 2	(b)	55.4%	76.9%	50.6%	70.1%
Experiment 3	(c)	54.3%	74.0%	49.3%	69.2%
Experiment 4	(b) (c)	57.1%	80.5%	52.4%	73.1%
Experiment 5	(b) (c) (d)	61.9%	83.3%	56.2%	77.0%
Experiment 6	(b) (c) (d) (e)	**62.5%**	83.0%	56.5%	76.8%
Experiment 7	(b) (c) (d) (f)	62.0%	82.9%	56.4%	76.7%
Experiment 8	(b) (c) (d) (e) (f)	62.4%	**83.4%**	**56.6%**	**77.2%**
Experiment 9	(b) (c) (d) (g)	58.8%	81.3%	53.9%	74.6%

Table 5: The results obtained from the analysis of the polarity detection method

4.2 Evaluation 2. Comparison of our affective lexicon *vs* Perez's lexicon

In the set of experiments for Evaluation 2, the SWePT method was evaluated comparing our affective lexicon and the Spanish lexicon by Perez [Perez-Rosas, 2012]. The experiments were carried out using the ten-fold cross validation technique, the SMO algorithm implemented with the *Weka* library and the corpus with 3100 comments.

The goal of this evaluation was to verify the behavior of the SWePT method when a different lexicon is used, in this case the Spanish lexicon by Perez [Perez-Rosas, 2012]. Therefore, the comparison was made using experiments 5 through 8 from evaluation 1 because they provide the best results. Tables 6, 7, and 8 show the results of the experiments. Many of the best results were obtained with our affective lexicon in Mexican Spanish, except for the results obtained in experiments 5, 6 and 7 with three categories (Table 8).

Besides, the results in five categories are better using the Affective lexicon in Mexican Spanish because our lexicon has more expressions to distinguish between the polarity categories, for example: intensifiers, interjections, negations and emoticons. Also, almost all the results obtained with the Spanish Lexicon by Perez are improved using both lexicons.

Experiments	Affective lexicon in Mexican Spanish					
	5 categories			3 categories		
	Precision	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**
Experiment 5	56.3%	56.2%	56.2%	76.8%	77.4%	77.0%
Experiment 6	56.6%	56.5%	56.5%	76.7%	77.3%	76.8%
Experiment 7	56.5%	56.4%	56.4%	76.6%	77.3%	76.7%
Experiment 8	56.6%	56.5%	**56.6%**	77.1%	77.7%	**77.2%**

Table 6: The results obtained with the Affective Lexicon in Mexican Spanish

Experiments	Spanish Lexicon by Perez					
	5 categories			3 categories		
	Precision	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**
Experiment 5	51.9%	48.1%	54.3%	75.9%	76.5%	76.1%
Experiment 6	52.1%	54.3%	54.2%	76.0%	76.6%	76.2%
Experiment 7	52.3%	54.7%	54.7%	76.1%	76.7%	**76.3%**
Experiment 8	56.5%	56.4%	**56.4%**	76.0%	76.6%	76.2%

Table 7: The results obtained with the Spanish Lexicon by Perez [Perez-Rosas, 2012]

Experiments	Both lexicons					
	5 categories			3 categories		
	Precision	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**
Experiment 5	52.5%	55.7%	55.6%	77.3%	77.9%	**77.5%**
Experiment 6	55.8%	57.1%	55.7%	76.8%	77.4%	77.0%
Experiment 7	56.8%	50.8%	**56.3%**	76.9%	77.5%	77.1%
Experiment 8	56.2%	56.6%	56.1%	76.7%	77.3%	76.9%

Table 8: The results obtained with our Affective lexicon in Mexican Spanish and the Spanish lexicon by Perez [Perez-Rosas, 2012]

4.3 Evaluation 3. Experiments with a specific-domain corpus

In the set of experiments for Evaluation 3, thee experiments were carried out using the specific-domain corpus created in section 3.5. The experiments were the following:

Experiment 1. We compared the manual labeling of the specific-domain corpus with the results obtained with the SWePT method. For example, the comment *the egg is too expensive* was labeled as negative by the people labeling. The same comment was taken as an input of the SWePT method (trained with the general corpus). Then, the result obtained with the SWePT method was compared with the manual labeling to verify whether the result was negative too. The comparison was registered in terms of precision, recall, and F-measure. Table 9 shows the results obtained. The average F-Measure of 43.1% indicates that almost half of the labels obtained by the SWePT method are the same labels when labeled manually.

	Precision	**Recall**	**F-Measure**
Positive	40.3%	32.0%	35.7%
Neutral	45.3%	51.8%	48.4%
Negative	42.9%	48.2%	45.4%
Average	42.8%	44.0%	**43.1%**

Table 9: The results of the comparison between the manual labeling and the SWePT method

Experiment 2. We compared the manual labeling of the specific-domain corpus with the results obtained with Wolfram Mathematica [Wolfram, 2016] instead of the SWePT method. We followed the same process as in Experiment 1. The recall in the neutral class was 69.9% and the precision in the neutral class was 27.7%. This shows that a great number of the comments were classified as neutral and that our method allowed for obtaining better results. Our method had an average F-Measure of 43.1% and the Wolfram Mathematica [Wolfram, 2016] had an average F-Measure of 29.6%. Table 10 shows the results.

	Precision	**Recall**	**F-Measure**
Positive	42.0%	35.4%	38.0%
Neutral	27.7%	69.9%	39.7%
Negative	46.0%	06.1%	10.85%
Average	38.6%	37.1%	**29.6%**

Table 10: The results of the comparison between the labeling manual and Wolfram Mathematica [Wolfram, 2016]

Experiment 3. We trained the SWePT method with the specific-domain corpus without pre-processing. A testing corpus was created to do this experiment. The testing corpus was composed by 500 comments of the same domain. The average F-Measure improved considerably (43.1%). The results are presented in Table 11.

	Precision	**Recall**	**F-Measure**
Positive	76.3%	77.0%	76.7%
Neutral	50.0%	45.1%	47.4%
Negative	65.5%	69.2%	67.3%
Average	65.9%	66.3%	**66.0%**

Table 11: The results of the evaluation of the SWePT method trained with a specific-domain corpus

Although, the size of corpus used in this experiments was the smallest, the results show that the method can be improved considerably by using a specific-domain corpus instead of a general corpus.

5 Case study: the Emotion-Bracelet interface

The SWePT web service was implemented with a physical interface called Emotion-Bracelet. This is a non-intrusive communication interface that is capable of expressing the emotional state of a user based on comments posted on Facebook. Since the Emotion-Bracelet is connected via Bluetooth with a Smartphone, the interface can obtain the comments posted by the user on Facebook, and the comments are analyzed by the web service in order to obtain the polarity of those comments [Molina, 2014].

The system works as follows: when a user posts a comment on her own Facebook account, the Emotion-Bracelet automatically extracts the post and calls the SWePT web service in order to identify the polarity of the comment. If the comment is *very positive* or *positive,* the Emotion-Bracelet shows the emoticon ☺. If the comment is *very negative* or *negative,* the Emotion-Bracelet shows the emoticon ☹. If the comment is neutral, the interface shows the emoticon 😐. In order to differentiate the polarity intensity, when the comment is *very positive,* the leds will light up green, when the comment is *positive,* the leds will light up yellow, when the comment is *very negative,* the leds will light up red and when the comment is *negative* the leds will light up orange [Molina, 2014].

Figure 3 shows an example of the performance of the Emotion-Bracelet where a user posts the following comment: "*you are one of the more beautiful pages that have been written in mi life <3 (in Spanish: tu eres una de las páginas más lindas que ha sido escrita en mi vida <3)*" on Facebook, the web service classifies the comment as happy, and the Emotion-Bracelet displays the emoticon ☺.

Figure 3: The Emotion-Bracelet is representing a happy emotional state

6 Conclusions and future work

The current work presents a hybrid approach for polarity detection, which combines a machine learning method with a lexicon-based method. According to the state of the art, machine learning methods give the best results; however, the main disadvantage is that methods of this kind require previous training and are domain dependent. The state of the art approaches demonstrate that the accuracy of a polarity detection system could be improved by combining different approaches. Generally, comments are very hard to deal with. This is mostly due to their size; sometimes they are too short, and therefore we are dealing with a lack of context.

In this paper, we have presented a manual labeled corpus with a topic-specific polarity; this kind of corpora can be useful to create a corpus for topic-based polarity models. The experiments also show that when there are more categories, more confusion can arise because it may be difficult to distinguish the degree of a category (e.g., *positive* and *very positive*). In general, the best results were obtained with only three categories.

The agreement among the annotators who manually labeled the corpus was evaluated with Krippendorff's alpha. The subjectivity involved in the interpretation and annotation of the comments had an impact on the results. The polarity is characterized by the context, personal experiences, prior knowledge, and people´s moods. Therefore, when people label a corpus, the interpretation of the emotional charge of the text can affect the results.

The SWePT method was applied in a case study using a physical interface called the Emotion Bracelet. This interface allows a user to express emotions (from their comments published in Facebook) in real time and in a non-intrusive way. The interface can be used in different contexts, such as emotional communication, entertainment, health, education, family care, etc.

We plan to do the following as future work. Since in many cases the polarity of a given word totally depends on the context in which it appears, lexicons based on a specific domain need be created with the affective word of specific domains. For instance, with the search string egg price, there are almost no associated words like *happy*, *sad, depressed*, etc.; instead, words like *expensive, poverty* and *dearth* appear constantly. First, the frequencies of words in a specific domain must be obtained. Then, the words with the highest frequency must be analyzed and then it must be determined whether or not those words with the highest frequencies are affective words. Finally, the words must be added to a lexicon that is based on the specific domain. Experiments to verify the viability of a lexicon based on a specific domain should also be performed.

The linguistic rules must be improved for better functioning of the feature extraction module and the verb tense should be taken into account. For example, in the sentence *I was sad, until I knew you*, the verb *was* is written in past continuous. This means that the person is not sad anymore. Even though this sentence is positive, the SWePT method would surely classify it as negative. Discourse analysis should be considered. For example, sometimes a positive or negative idea is expressed at the beginning of a text, but the text ends with an idea with the opposite polarity, which change the overall polarity of the text. For example, the sentence *We did our work the best what we could, but it was not enough* starts with a positive phrase, but ends with a negative one, and the overall text is negative.

Also, some experiments to identify irony should be implemented. Irony is a very common issue that is present in opinions and represents an obstacle because irony disguises the true attitudes expressed in the text [Hernandez, 2015]. This fact has a direct impact on the accuracy of any Opinion Mining system [Reyes, 2013; Ghosh, 2015]. Therefore, problems of this kind must be addressed.

Acknowledgements

This work has been partially supported by the Sectorial Fund CONACyT-INEGI: Project with ref. 208471, INFOTEC, Mexico. And, also by the project CNDT-PYR-2015-0016, CENIDET, Mexico. The work of the third author was in the framework of the SomEMBED MINECO TIN2015-71147-C2-1-P research project. The National Council for Science and Technology (CONACyT Mexico) has funded the research work of Delia Irazu Hernandez Farias (Grant No. 218109/313683 CVU-369616).

References

[Baca-Gomez, 2014] Baca-Gomez, Y. R.: Development of a Web Service for determining the Polarity of comments from Social Network in Spanish, National Center for Research and Technology Development, Cuernavaca, Morelos, Mexico, Master Thesis, 2014, https://www.dropbox.com/s/j9c4m22135v573a/Resources.rar?dl=0

[Baccianella, 2010] Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. Proceedings of the 7th International Conference on LREC'10, 2200-2204, La Valletta, Malta, May 2010.

[Balage-Filho, 2014] Balage-Filho, P.P, Pardo, T. A. S.: NILC_USP: A Hybrid System for Sentiment Analysis in Twitter Messages, Proceedings of the 8th International Workshop on Semantic Evaluation, 428-432, Dublin, Ireland, August 2014.

[Balahur, 2012] Balahur, A., Hermida, J. M., Montoyo, A.: Detecting implicit expressions of emotion in text: A comparative analysis, decision support systems, computational approaches to subjectivity and Sentiment Analysis and service science in information systems research - special issue on PACIS 2010, Volume 53, Issue 4, 742-753, November 2012.

[Benamara, 2007] Benamara, F., Cesarano, C., Picariello, A., Recupero, D. R., Subrahmanian, V. S.: Sentiment Analysis: adjectives and adverbs are better than adjectives alone, Proceedings of the International Conference on Weblogs and Social Media, 2007.

[Bespalov, 2011] Bespalov, D., Bai, B., Qi, Y., Shokoufandeh, A.: Sentiment Classification Based on Supervised Latent n-gram Analysis. Proceedings of the 20th ACM International Conference on Information and Knowledge Management, 375-382, Glasgow, United Kingdom, October 2011.

[Cruz, 2008] Cruz, F. L., Troyano, J. A., Enriquez, F., Ortega J.: Experiments in sentiment classification of movie reviews in Spanish, Procesamiento del Lenguaje Natural, No. 41, 73-80, Junio 2008.

[Diaz-Rangel, 2013] Diaz-Rangel, I., Suarez-Guerra, S., Sidorov, G.: Detection of affection in texts in Spanish based on the linguistic context for speech synthesis *(in Spanish)*, The *Center* for *Computing Research* of the *National Polytechnic Institute. PhD Thesis, 2013.*

[Fernandez-Anta, 2013] Fernandez-Anta, A., Nuñez-Chiroque, L., Morere, P., Santos, A.: Sentiment Analysis and Topic Detection of Spanish Tweets: A Comparative Study of NLP techniques, Natural Language Processing, No. 50, 45-52, March 2013.

[Fleiss, 2013] Fleiss, J. L., Levin B., Paik, M.: Statistical Methods for Rates and Proportions, Third Edition, Wiley Series in Probability and Statistics, John Wiley & Sons Co., 2003.

[Garcia-Diaz, 2006] Garcia-Diaz, E. E., Rangel, P. A., Lozano-Martinez, F., Adaptive Support Vector Machines for time series prediction, XI Symposium signal processing, images and artificial vision 2006.

[Gezici, 2013] Gezici, G., Dehkharghani, R., Yanikoglu, B., Tapucu, D., Saygin, Y.: SU-Sentilab: A Classification System for Sentiment Analysis in Twitter, 2nd Joint Conference on Lexical and Computational Semantics, Volume 2: 7th International Workshop on Semantic Evaluation, 471-477, Atlanta, Georgia, June 2013.

[Ghosh, 2015] Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Barnden, J., Reyes, A.: Semeval-2015 task 11: Sentiment Analysis of figurative language in Twitter. Proc. 9th Int. Workshop on Semantic Evaluation (SemEval 2015), Co-located with NAACL, 470–478, Denver, Colorado, June 2015.

[Glorot, 2011] Glorot, X., Bordes, A., Bengio, Y.: Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach, Proceedings of the 28th International Conference on Machine Learning, 513-520, Bellevue, WA, USA, 2011.

[Hernández, 2015] Hernández, I., Benedí, J. M., Rosso, P.: Applying basic features from Sentiment Analysis on automatic irony detection. In: Proc. 7th Iberian Conf. on Pattern Recognition and Image Analysis (ibPRIA), Pattern Recognition and Image Analysis, Springer-Verlag, LNCS, Volume 9117, 337-344, Santiago de Compostela, Spain, June 2015.

[Hall, 2009] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H.: The WEKA data mining software: an update, ACM SIGKDD Explorations Newsletter, Volume 11, Issue 1, 10-18, June 2009.

[Krippendorff, 2011] Krippendorff, K.: Computing Krippendorff's Alpha-Reliability: University of Pennsylvania, Annenberg School for Communication, 2011.

[Letouze, 2014] Letouze, E., Jütting, J.: Official Statistics, Big Data and Human Development: Towards a New Conceptual and Operational Approach, Data-Pop Alliance white papers series, Partnership in Statistics for Development in the 21st Century (PARIS21), December 2014.

[Liu, 2012] Liu, B.: Sentiment Analysis and Opinion Mining, Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, May 2012.

[Marchand, 2012] Marchand, M., Ginsca, A. L., Besançon R., Mesnard, O.: [LVIC-LIMSI]: Using Syntactic Features and Multi-polarity Words for Sentiment Analysis in Twitter, 2nd Joint Conference on Lexical and Computational Semantics, Volume 2: Proceedings of the 7th International Workshop on Semantic Evaluation, 418-424, Atlanta, Georgia, June 2013.

[Martinez-Camara, 2011] Martinez-Camara, E., Martin-Valdivia, M. T., Perea-Ortega, J. M., Ureña-López, L. A.: Opinion Classification techniques applied to a Spanish corpus, Natural Language Processing, No. 47, 163-170, September 2011.

[Martinez-Camara, 2011b] Martinez-Camara, E., Martin-Valdivia, M.T., Ureña-Lopez, L. A.: Opinion Classification Techniques Applied to a Spanish Corpus, Natural Language Processing and Information Systems, LNCS, Volume 6716, 169-176, 2011.

[Martin-Valdivia, 2013] Martin-Valdivia, M. T., Martinez-Camara, E., Perea-Ortega, J. M., Ureña-Lopez, L. A.: Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches, Volume 40, Issue 10, 3934-3942, August 2013.

[Martin-Wanton, 2010] Martin-Wanton, T., Pons-Porrata A.: Opinion Polarity Detection Using Word Sense Disambiguation to Determine the Polarity of Opinions, Proceedings of the International Conference on Agents and Artificial Intelligence, Volume 1 – Artificial Intelligence, Valencia, Spain, January 2010.

[Molina, 2014] Molina, A., Martinez, A., Sanchez, W.: Emotion-Bracelet: An emotional interface, National Meeting of Computer Science, Mexican Society of Computer Science, November 2014.

[Morgan, 1988] Morgan Rick L., Heise David: Structure of Emotions. Social Psychology Quarterly, Volume 5 No. 1, 19–31, 1988.

[Ortigosa-Hernandez, 2012] Ortigosa-Hernandez, J., Rodriguez, J. D., Alzate L., Lucania M., Inza I., Lozano, J. A.: Approaching Sentiment Analysis by using semi-supervised learning of multi-dimensional classifiers, Neurocomputing, Data Mining Applications and Case Study, Volume 92, 98-115, September 2012.

[Padro, 2012] Padro, L., Stanilovsky, E.: FreeLing 3.0: Towards Wider Multilinguality, Proceedings of the Eighth International Conference on LREC'12, European Language Resources Association, Istanbul, Turkey, May 2012.

[Palanisamy, 2013] Palanisamy, P., Yadav, V., Elchuri, H.: Serendio: A simple and Practical lexicon-based approach to Sentiment Analysis, 2nd Joint Conference on Lexical Computational Semantics, Volume 2: Seventh International Workshop on Semantic Evaluation, 543-548, Atlanta, Georgia, June 2013.

[Pang, 2002] Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques, Proceedings of the ACL-02 conference on Empirical methods in Natural Language Processing, Association for Computational Processing, Volume 10, 79-86, 2002.

[Perez-Rosas, 2012] Perez-Rosas, V., Banea, C., Mihalcea, R.: Learning Sentiment Lexicons in Spanish, Proceedings of the Eighth International Conference on LREC'12, European Language Resources Association, Istanbul, Turkey, May 2012.

[Platt, 1998] Platt, J. C.: Fast Training of Support Vector Machines Using Sequential Minimal Optimization, Microsoft Research, Proceedings Advances in Kernel Methods – Support Vector Learning, MIT Press, January 1998.

[Poria, 2014] Poria, S., Cambria, E., Winterstein, G., Huang, G. B.: Sentic patterns: Dependency-based rules for concept-level Sentiment Analysis, Knowledge-Based Systems, Volume 69, 45-63 Elsevier B. V. October 2014.

[Reyes, 2013] Reyes, A., Rosso, P.: On the Difficulty of Automatically Detecting Irony: Beyond a Simple Case of Negation. In: Knowledge and Information Systems, Volume 40, No. 3, 595-614, September 2013.

[Russell, 1980] Russell, J. A.: A Circumplex Model of Affect, Journal of Personality and Social Psychology. Volume 39, No. 6, 1161–1178, University of British Columbia, Vancouver, Canada,1980.

[Rushdi, 2011] Rushdi Saleh, M., Martin-Valdivia, M. T., Montejo-Raez, A., Ureña-López, L. A.: Experiments with SVM to classify opinions in different domains, Expert Systems with Applications, Volume 38, Issue 12, 14799-14804, November-December 2011.

[Sacharin, 2012] Sacharin, V., Schlegel K., Scherer K. R.: Geneva Emotion Wheel rating study (Report). Geneva, Switzerland, University of Geneva, Swiss Center for Affective Sciences, 2012.

[San Vicente, 2014] San Vicente Roncal, I., Saralegi Urizar, X.: Looking for Features for Supervised Tweet Polarity Classification, Proceedings on TASS 2014 – Workshop on Sentiment Analysis at SEPLN, Girona, Spain, September 2014.

[Saralegi, 2012] Saralegi, X., San Vicente I., TASS: Detecting sentiments in Spanish tweets, Proceedings of the TASS Workshop at SEPLN, 2012.

[Sarmento, 2009] Sarmento, L., Carvalho, P., Silva, M. J., de Oliveira, E.: Automatic creation of a reference corpus for political Opinion Mining in user-generated content, Proceedings of the 1st International CIKM workshop on Topic-Sentiment Analysis for mass opinion, 29-36, New York, 2009.

[Scherer, 2013] Scherer, K. R., Shuman V., Fontaine J. R. J., Soriano C.: The GRID meets the Wheel: Assessing emotional feeling via self-report. Components of emotional meaning: A sourcebook. Oxford, Oxford University Press, 2013.

[Scherer, 2005] Scherer K. R.: What are emotions? And how can they be measured? Trends and developments: research on emotions. Social Science Information, SAGE Publications, Volume 4, No. 4, 695–729, 2005.

[Schulz, 2010] Schulz, J. M., Womser-Hacker C., Mandl, T.: Multilingual Corpus Development for Opinion Mining, Proceedings of the Seventh International Conference on LREC'10, European Language Resources Association, Valletta, Malta, May 2010.

[Stone, 1966] Stone P. J., Dunphy D. C., Smith M. S., Ogilvie D. M.: The General Inquirer: A Computer Approach to Content Analysis. MIT Press, Oxford, England, 1966.

[Strapparava, 2004] Strapparava C., Valitutti A.: WordNet-Affect: An Affective Extension of WordNet. Proceedings of the 4th International Conference on LREC'04, 1083–1086, ITC-irst, Instituto per la Ricerca Scientifica e Tecnologica, Povo Trento, Italy, 2004.

[Taboada, 2011] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-Based Methods for Sentiment Analysis, Computational Linguistics, Volume 37, No. 2, 267-307, May 2011.

[Tang, 2015] Tang, D., Qin, B., Liu T.: Deep learning for sentiment analysis: successful approaches and future challenges. WIREs Data Mining Knowl Discov, No. 5, 292-303, 2015.

[Turney, 2002] Turney, P. D.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, Pennsylvania, 417-424, July 2002.

[Vilares, 2013] Vilares, D., Alonso, M. A., Gomez-Rodriguez, C.: Supervised polarity classification of Spanish tweets based on a linguistic knowledge, Proceedings of the 2013 ACM symposium on Document engineering, 169-172, September 2013.

[Wang, 2014] Wang, G., Sun, J., Ma, J., Xu, K., Gu., J.: Sentiment classification: The contribution of ensemble learning, Decision Support Systems, Volume 57, 77-93, January 2014.

[Wiebe, 2006] Wiebe, J., Wilson, T., Cardie, C.: Annotating Expressions of Opinions and Emotions in Language, Proceedings of the 5th International Conference on LREC'06, Volume 39, Issue 2-3, 165-210, February 2006.

[Wilson, 2005] Wilson, T., Hoffman P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., Patwardhan, S.: OpinionFinder: A system for subjectivity analysis. Proceedings of HLT/EMNLP Interactive Demonstrations, 34–35, Vancouver, 2005.

[Wolfram, 2016] Wolfram Mathematica, 2016, https://www.wolfram.com/mathematica/

[Xia, 2011] Xia, R., Zong, C., Li, S.: Ensemble of feature sets and classification algorithms for sentiment classification, Information Sciences, Volume 181, Issue 6, 1138-1152, March 2011.