

Unsupervised Opinion Polarity Detection based on New Lexical Resources

Mario Amores, Leticia Arco, Claudia Borroto

Universidad Central "Marta Abreu" de Las Villas, Department of Computer Science,
Cuba

marioamores@uclv.cu, leticiaa@uclv.edu.cu

Abstract. There are polarity detection techniques based on the lexicon of opinion words and those based on machine learning techniques. In this paper, we focus on unsupervised polarity detection using lexical resources. We present the SentiWordNet 4.0 and the SpanishSentiWordNet in order to solve the detected drawbacks of previous resources. The integration of the proposed resources is solved by combining them in the PolarityDetection library, which is integrated to PosNeg Opinion 2.0 and facilitates obtaining high accuracy and recall values.

Keywords. Unsupervised polarity detection, lexical resources, opinion mining.

1 Introduction

Sentiment analysis is a challenging task in Natural Language Processing (NLP). Due to its tremendous value for practical applications, it has received a lot of attention, and it is perhaps one of the most widely studied topics in the NLP field nowadays [1, 2].

Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space [3].

The opinion conveyed by a text can be expressed by very subtle and varied words, therefore it is often difficult to exactly determine it. The classification of sentiments (polarity) is a sub-task in opinion detection [4]. It consists in determining whether an opinion in a given document is positive or negative, which has been challenged at the Text Retrieval Conference

(TREC) Blog Track since 2006. The approaches explored by track participants can be devised in two types of approaches for opinion and polarity detection. Some of them are based on a lexicon of opinion words [6, 7, 8], others, on machine learning techniques [9, 10, 11]. The third type of works use a mixed approach (machine learning and lexicon) [12].

Most of the works on polarity detection use the first approach, which applies a lexicon of opinion words and is usually unsupervised [6, 13, 14]; this is the approach we follow in this paper.

The lexicon can be general (such as SentiWordNet [15], WordNet Affect [16], General Inquirer [17]), built manually or generated automatically from the corpus (words that contain an opinion are taken directly from the corpus). Each word in the lexicon is associated with opinion and polarity scores. These scores are exploited by different approaches to compute the opinion (or polarity) score of a document. A simple method is to assign a score equal to the total number of words containing an opinion (or polarity) in the document [18, 19].

Unfortunately, lexical resources currently used in opinion mining sometimes assign wrong scores to words, they are all very scattered, and some of them use old formats. Therefore, in some cases it becomes difficult to integrate these tools in one application. Thus, new resources are needed to report better results.

On the other hand, although most of the research conducted focus on English texts [20, 21, 22], the number of papers on the treatment of other languages is increasing every day [6, 23, 24, 25]. Thus, another disadvantage of lexical resources is that they are generally focused only on the English

language; then, new lexical resources are needed for other languages. We are particularly interested in the English and Spanish languages.

In this paper, we focus on unsupervised opinion detection using lexical resources. We present two new lexical resources: SentiWordNet 4.0 and SpanishSentiWordNet, in order to solve the drawbacks of lexical resources raised above. Also, a new Intralinguistic Index is presented for assisting the development of the SpanishSentiWordNet. The integration of the proposed resources is solved by combining them in the new PolarityDetection library. Finally, in order to show the applicability of our resources, PosNeg Opinion application uses them and several experiments have been performed in order to analyze major issues in our approach and to compare it with other approaches.

The paper is structured as follows: in section 2 related works are discussed, in section 3 the new lexical resources are described. In section 4 the polarity values assigned to some terms in the lexical resources are illustrated. Section 5 explains how to use the developed resources in polarity detection. In sections 6 and 7 experimental results are displayed. Finally, section 8 is the conclusion of the paper.

2 Related Works

There are a lot of papers about polarity detection of opinions [6, 7, 8]. In this section we will focus on some approaches specifically developed for the English [4, 7, 8] or Spanish languages [6, 10, 21] or those approaches that are not language dependent [14, 26, 27]. Spanish and English are the languages we are interested in our analysis.

The system proposed in [9] is based on a robust parser that provides information to feed Support Vector Machine (SVM) classifiers with linguistic features dedicated to aspect categories and aspect categories polarity classification. This combination proved to be an interesting platform to implement a category/polarity detection system. However, it is necessary to improve performances on the neutral and conflict polarities, which rely less on specific words than on a more global interpretation of the content.

Another aspect based opinion detection system using a robust deep syntactic parser [31] was proposed in [9], from which semantic relations of opinion were calculated. The polarity lexicon was built using existing resources and also by applying classification techniques over large corpora, while the semantic extraction rules were hand-crafted. Detecting the aspect terms and aspect categories and their corresponding polarities are the principal tasks in this approach. One classifier is trained for detecting the categories and further, for each category, a separate classifier is trained for detecting the polarities corresponding to that particular category. The approaches proposed in [9] and [32] are supervised, so they suffer from the disadvantages associated with supervised methods: the need for a training corpus and dependence on the domain where the model was obtained.

A novel approach to sentiment polarity classification in Twitter posts by extracting a vector of weighted nodes from the graph of WordNet [29] is presented in [8]. These weights are used in SentiWordNet to compute a final estimation of the polarity. Therefore, the method proposes a non-supervised solution that is domain-independent. One of the challenges when processing tweets is their short length, therefore in most cases there are few elements of information to decide whether the texts are positive or negative. This obstacle is more evident when it is followed by an unsupervised approach because it depends largely on the terms presented in a text and context. Therefore, the proposed method intends to expand the few concepts that are in tweets in order to calculate the global polarity of the tweet using Personalized Page Rank vectors (PPV) and taking the graph of WordNet, where nodes are synsets and axes are the different semantic relations between them. By combining a random walk algorithm that weights synsets from the text with polarity scores provided by SentiWordNet, it is possible to build a system comparable to an SVM based supervised approach in terms of performance. A pending task in this approach is how to deal with negation, as the score from SentiWordNet should be considered in a different way in the final computation if the original term comes from a negated phrase. Another disadvantage is that the proposed solution needs two main resources: a

graph to connect terms and polarity weights for individual terms. Maybe, it is not possible to construct a graph for a specific controlled domain.

In [9] an unsupervised opinion mining system was proposed to determine the polarity of sentences; i.e. to classify the sentences as positive, negative, or neutral. Among the various available methods in the unsupervised technique, the dictionary-based approach was used to determine the orientation of sentences. WordNet [29] was used as a dictionary to determine the opinion of words and their synonyms and antonyms. The system, similar to the approach proposed in [33], generates clusters of positive, negative, and neutral sentences, which will be easier for users to read and analyze, and which will help them in taking the decision whether the product is to be purchased or not. The polarity of a given sentence is determined on the basis of the majority of opinion words. Part-Of-Speech (POS) tagger [34, 35] is necessary to identify the opinion words. Synonyms and antonyms are determined with the help of WordNet. Polarity is determined on the basis of majority of opinion words: if the number of positive words is bigger, then the sentence is positive, otherwise negative, and if the number of positive and negative words are equal, then the sentence shows the neutral polarity. Negation is also handled in the system. If the opinion word is preceded by the negative particle "not", then the polarity of that sentence is reversed. This approach does not determine the polarity based on the aspects of sentences and it has to improve the analysis of sentences containing some clauses, such as *neither-nor* and *either-or*.

Until now, most of the published methods for polarity classification have been applied to English texts, but other languages are becoming increasingly important [6, 23, 24, 25]. A new resource for the Spanish sentiment analysis research community is presented in [6]. The authors generated a Spanish lexicon, the Spanish Opinion Lexicon (SOL), based on one of the most widely-used English lexicon, the Bing Liu English Lexicon (BLEL) [21]. Specifically, they focus on the use of opinion words. The manually reviewed lexicon is improved in order to obtain the final list of words named iSOL (improved SOL) and eSOL (enriched SOL) and it was developed for integrating external knowledge. A lexicon such as

iSOL and eSOL can be used as the sole semantic resource or it can be used as another element within the workflow of a polarity classification system. The developed lexicons are freely available and they are valuable resources for the Spanish research community.

A novel approach in the generation of knowledge resources for Sentiment Analysis by crawling the vast flow of micro-texts published in social media every second was introduced in [37]. By filtering a small (but yet a huge) part of these streams and categorizing them semi-automatically, they produce a resource for Polarity Classification with little human intervention. Thus, the idea behind it is to represent each tweet to be classified as a ranked vector of feelings. Then, a final polarity value is calculated from this vector. Many issues remain open, like lexical normalization, so informal expressions could be better conflated, jargon properly represented, and emoticons also considered. The obtained results are not the best compared to those obtained on the same corpora, and one of the reasons is that this is an unsupervised approach. Besides, the language independence shown by the method is interesting in those domains where resources in certain languages are not always available. Another main drawback is the manual annotation of most frequent emotions extracted in order to create the index. Anyhow, additional resources could be used here, like translated versions of SentiWordNet or other concept-based indexes, like SenticNet [38].

A set of experiments classifying Spanish tweets according to sentiment and topic is presented in [10]. Here the use of stemmers and lemmatizers, n-grams, word types, negations, valence shifters, link processing, search engines, special Twitter semantics (hashtags), and different classification methods are evaluated. The authors came to three important conclusions [10]: none of the techniques explored is the silver bullet for Spanish tweet classification, none made a clear difference when introduced in the algorithm; tweets are very hard to deal with, mostly due to their brevity and lack of context; and there is still a lot of room for improvement, justifying further efforts.

As we can see in the above approaches, a lot of solutions use resources for polarity detection: taggers (e.g., TreeTagger [28]), dictionaries (e.g., WordNet [29]), concept-based indexes (e.g.,

SenticNet [38]), and specific sentiment or polarity tools (e.g., SentiWordNet [15]). However, some of these resources have disadvantages, therefore, they affect the quality of the final result to be obtained; for instance, the accuracy and recall values obtained in [10, 37] are lower than 0.5 and 0.7, respectively. Therefore, to work on creating or enriching existing resources should be useful to improve the performance of polarity detection systems.

3 SentiWordNet 4.0 and SpanishSentiWordNet

Four different versions of SentiWordNet have been discussed in publications [14]: SentiWordNet 1.0 [15], SentiWordNet 1.1 [39], SentiWordNet 2.0 [40], and SentiWordNet 3.0 [41]. Each version has tried to eliminate the disadvantages of the predecessors. In this paper we focus on the disadvantages of the SentiWordNet 3.0 and the necessity of a resource for Spanish, in order to create new lexical resources improving, on the one hand, the format and the annotation of terms, and on the other hand, considering the Spanish language. In addition, for creating the SpanishSentiWordNet it is important to use the intralinguistic index [42]; for this reason, a new intralinguistic index for the SpanishSentiWordNet is presented in this paper.

3.1 SentiWordNet 4.0

In this paper, we introduce a new version of the SentiWordNet, named SentiWordNet 4.0. This version solves the disadvantages presented in SentiWordNet 3.0 concerning the format and polarity of terms. Thus, we will briefly describe the general characteristics of SentiWordNet 3.0, useful for explaining the new modifications later.

The entry format of the SentiWordNet 3.0 consists of a term (a term can be a word or a list of words) followed by a Tab character, the POS label, a space, the positive value of the term, another space, and the negative value of the term, as we show in Figure 1. The POS labels can be n, a, v, or r; where n means noun, a means adjective, v means verb, and r means adverb.

As we can see in Figure 1, some terms consist of more than one word and characters _ and -, these are used for written phrases and compound words, respectively. The textual analysis considering phrases and compound words is more useful in other applications of NLP than in polarity detection, because in most of the cases, it does not provide additional information of individual words, and conversely, it significantly complicates the process of having to look for pairs, triples, and quads of words. As it can be seen circled in Fig. 1, several terms receive positive and negative polarity values equal to 0. This is a big problem for polarity detection applications which use SentiWordNet 3.0, because the main service that should provide this resource is to offer real polarity values for polarity classification systems.

SentiWordNet 3.0 was created by a mixture of linguistic techniques and statistical classifiers. Its construction was semiautomatic, this means that the results were not manually verified; for that reason, some classifications may be incorrect. For instance, the synset#1 of the noun *flu* (*{influenza, flu, grippe}* - (*an acute febrile highly contagious viral disease*)) was classified as Positive=0.75 and Negative=0.0, without taking into account that this synset consists of a lot of negative words.

Considering the above disadvantage, the first modification that we introduce in SentiWordNet 4.0 is re-labelling the polarity values of each term. To do this, an analysis of matches between the positive and negative terms of SentiWordNet with two lists of positive and negative words published in [33] was performed. We consider a term positive or negative in SentiWordNet 3.0 if the highest score associated to this term is positive or negative, respectively. The reference lists of words have 2005 positive words and 4781 negative

```

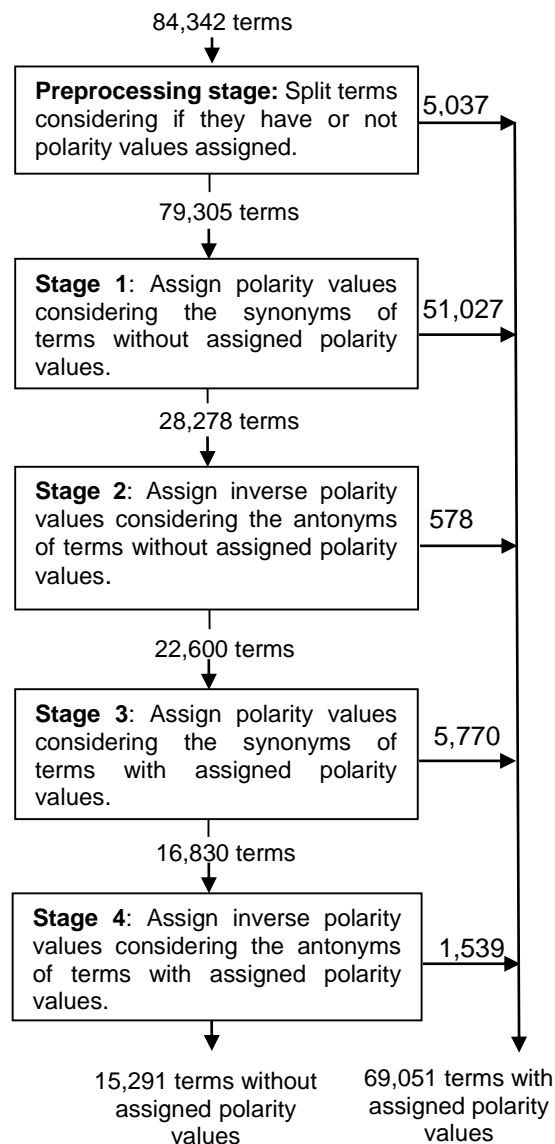
able a 0.125 0
unable a 0 0.75
adductive adducting adducent a 0.0 0.0
parturient a 0.25 0.0
surface-assimilative adsorptive adsorbent a 0.0 0.0
torrential a 0.0 0.375
ideal a 0.0 0.0
ready_to_handy a 0.125 0.125

```

Fig. 1. The format of SentiWordNet 3.0

Table 1. Positive and negative matches between the reference lists and SentiWordNet 3.0

Terms in SWN	Positive matches	Negative matches	Total
117659	735	1175	1910

**Fig. 2.** General scheme for assigning polarity values

words belonging to diverse domains and contexts [33]. The positive and negative matches between the reference lists and SentiWordNet 3.0 are shown in Table 1.

As we can see in Table 1, only 1910 terms have the same polarity classification in the SentiWordNet and the reference lists. These results show how deficient are the scores in the SentiWordNet 3.0, because the polarity values are inaccurate. Thus, in this research we improve the assignment of positive and negative score values to each term and consequently we improve the matches between SentiWordNet and the reference lists.

As we stated before, in some cases more than one word per term is contained in the SentiWordNet 3.0 format; this means that they use some synonyms to identify a specific term. Our proposal for solving this problem in SentiWordNet 4.0 is to create a new term corresponding to each word which describes a specific term in the SentiWordNet 3.0. Each new term will receive the same label and polarity values.

In addition, as mentioned above, the presence of compound words and phrases complicates the analysis and provides no benefits. Thus, we eliminate the signs `_` and `-` in the original compound words and phrases. After that, we extracted the simple terms and we deleted the terms classified as stopwords, they are terms that offer no meaning in the English language. Then, the possible labels for each term were sought by the `getPos` function in WordNet [29]. Thus, we include in SentiWordNet 4.0 the new terms with their labels and the positive and negative polarity values.

This disaggregation process of compound words and phrases caused that simple terms already existed previously in the SentiWordNet 3.0 were generated. For that reason, repetitions of the same term with the same associated label were eliminated. For example, the term *ideal* appeared five times, two times as adjective and three times as noun. In this case, if the original term had polarity values greater than zero, we kept these values, otherwise, the polarity associated with the same term incorporated later was assigned. Besides, eliminating repeated terms also excluded those terms that were proper names, as these do not affect the opinion polarity. Thus, we solved the

first detected disadvantage and we re-labelled the polarity values of each term. SentiWordNet 4.0 is formed by 9030 words tagged as verbs, 53599 words tagged as nouns, 18076 as adjectives, and 3637 words tagged as adverbs, for a total of 84342 tagged terms.

Another deficiency found in the SentiWordNet 3.0 tool is that most of the terms it comprises have zero positive and negative polarity values. Thus, 80795 terms have positive and negative polarity values equal to 0, representing the 95.8% of the total number of tagged terms. This phenomenon limits this resource considerably, since its main objective is the classification of the polarity of opinions. If a resource does not have enough words tagged correctly, it is not possible to exactly classify the polarity of opinions.

To solve the problem posed above, we search the positive and negative terms from the SentiWordNet 3.0 in the reference lists [33], in order to add 1 to the positive value if the term appears in the reference list of positive words and to add 1 to the negative value if the term appears in the reference list of negative words.

So far, there have only been granted positive or negative polarity to 5037 terms in SentiWordNet 4.0; thus, the 95.4% of the total of terms does not have the polarity values assigned. Five stages were defined for increasing the tagged terms in SentiWordNet 4.0. In Figure 2 a general schema is shown, which represents the defined procedure for increasing the number of tagged terms considering the five stages. The starting point of our procedure is the list of 84342 terms that we included in SentiWordNet 4.0. The preprocessing stage only split terms considering if they have polarity values assigned or not. Using WordNet is extremely important for the rest of the stages.

In Stage 1, for each term without assigned polarity values, we found its synonyms in WordNet using the function `getAllSynonyms`. The idea here is to search in the list of terms without assigned polarity values those that coincide with a found synonym and assign the corresponding polarity values. This stage was very useful, because we could assign polarity values to 51027 terms. After finishing this stage, only 34.5% of terms remained without polarity values.

Stages 3 and 4 are similar to stages 1 and 2, respectively. The difference is that in stages 1 and

2 we look for synonyms and antonyms of terms without polarity values, respectively, and in the stages 3 and 4, we look for synonyms and antonyms of the terms that already have polarity values assigned. As a result of Stage 3, it was observed that the number of terms with polarity values was increased by 5770 terms; also as a result of Stage 4, 1539 other terms received their polarity values.

As a result of all stages, we increased the number of terms with polarity values to 69051, representing the 81.9% of terms that we included in SentiWordNet 4.0. Unfortunately, we could not assign polarity values to 15991 terms.

3.2 Intralinguistic Index

The Intralinguistic Index (ILI) is an unstructured base of concepts, which aims to provide an effective mapping among different languages. Each concept is represented as a record of ILI representing a reference to the associated synset source [42]

The format of an entry in ILI includes the Spanish term, a space character, the POS label of the term, a Tab character, and several identifiers of

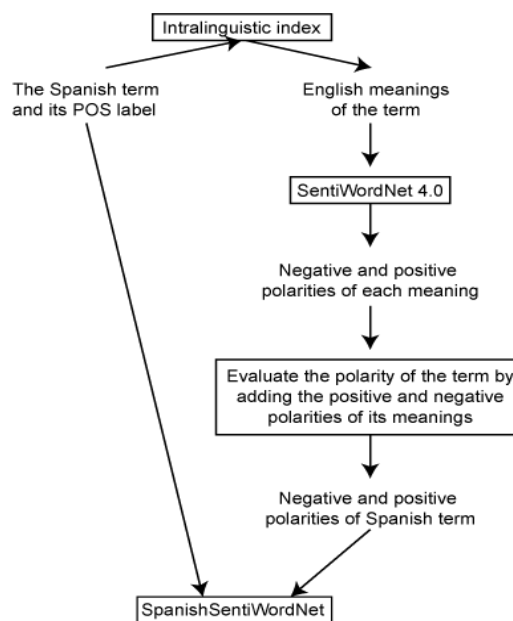


Fig. 3. Process for including a term in our SpanishSentiWordNet

different semantic relationships of the term, another Tab character, and finally, the English meanings of the term separated by space characters. As an example consider the term *agresor*: *agresor* n 09195176 09158637 09848308 *attacker assailant aggressor assaulter aggressor robber*.

The intralinguistic indices emerged as a list of synsets in WordNet 1.5, but they have been adapted to relate different language WordNets [29]. Further, this resource was used in the creation of EuroWordNet [43].

We detected two disadvantages in the Intralinguistic Index: some fields are not useful for opinion mining and some English meanings are repeated. These errors make the use of this resource slower with a performance which does not report good results. To solve the above mentioned disadvantages, the Intralinguistic Index format was modified. The identifiers of different semantic relationships and repeated English meanings were deleted. In the new format, the above example will be presented as follows:

agresor n *attacker assailant aggressor assaulter robber*.

3.3 SpanishSentiWordNet

To create the SpanishSentiWordNet, it was important the use of previously modified resources: the new Intralinguistic Index of WordNet for the Spanish language and the newly created SentiWordNet 4.0. We created the SpanishSentiWordNet using the above mentioned resources and following the method proposed in [14] which suggests to “evaluate the polarity of a term by adding the positive and negative polarities of its meanings”.

For each Spanish term included in the Intralinguistic Index, we follow the process described in Figure 3 for creating the SpanishSentiWordNet. The Intralinguistic Index offers us each Spanish term and its POS label; thus, we can include this information directly in our SpanishSentiWordNet. Then, only the positive and negative polarity values are lacking. We use the

Table 2. The polarity values of the term *kill* in SentiWordNet 3.0

POS	Positive polarity values	Negative polarity values	Sense
n	0	0	kill#1
n	0	0	kill#2
v	0	0.5	kill#1
v	0	0	kill#2
v	0	0	kill#3
v	0	0.125	kill#4
v	0.25	0.375	kill#5
v	0.75	0.125	kill#6
v	0	0	kill#7
v	0	0	kill#8
v	0	0	kill#9
v	0	0	kill#10
v	0.375	0	kill#11
v	0	0	kill#12
v	0.25	0.75	kill#13
v	0	0	kill#14
v	0.5	0.25	kill#15

Intralinguistic Index and the SentiWordNet 4.0 for obtaining the polarity values. First, we extract all English meanings of the term from the Intralinguistic Index. Second, we search in our SentiWordNet 4.0 and extract the negative and positive polarities of each meaning. Third, we compute the polarity of each term by adding the positive and negative polarities of its meanings following the method proposed in [14]; however, other aggregation operators could be used for combining the meaning polarity values. Thus, we have the whole information for including the Spanish terms, their POS labels, and their positive and negative polarity values in the SpanishSentiWordNet 4.0.

Finally, 43525 terms were included in the SpanishSentiWordNet; of these, 30289 nouns, 8664 adjectives, and 4572 verbs. The SpanishSentiWordNet format consists of a Spanish term, a Tab character, its POS label followed by a space, its positive polarity value, another space, and its negative polarity value.

4 Illustrating the Polarity Values of Some Terms in Lexical Resources

In this section we show the changes of the polarity values of some terms from SentiWordNet 3.0 to SentiWordNet 4.0. The term *kill* allows us to illustrate how the polarity values change from one lexical resource to another. Table 2 shows the polarity values of the term *kill* in SentiWordNet 3.0.

The term *kill* indicates negative polarity in most sentences except in some contexts, for example, "kill laughter" and "I enjoyed the film *Time to Kill*". Therefore, only senses 1, 4, and 6 are properly rated. Most senses with positive and negative polarity values equal to 0 are incorrectly rated. Sense 5 is defined as a negative term, but its positive polarity value is greater than zero; however, the definition of that sense is *be the source of great pain for*, so it is entirely negative. The same applies to sense 13 whose definition is *tire out completely*. Sense 11 is rated as an entirely positive term; however, its definition *drink down entirely* can be negative, for example, *She killed a bottle of brandy that night*. Something similar happens with sense 15 whose definition is *destroy a vitally essential quality of or in*, although this

sense has a negative score of the term *kill*, it is not enough to make it a negative term.

Now we will show the transformations made to the scores of the term *kill* in order to obtain its final scores in the SentiWordNet 4.0. Some authors argue that the polarity term classification will be better if we make an analysis for each sense. However, this approach does not provide good results when we work with short texts such as opinions or informal texts from social networks. Therefore, in these cases, an analysis by senses creates noise in the polarity opinion classification. Hence, first we decided to add the scores of all senses for each specific POS label. Table 3 shows the obtained polarity values. Positive and negative polarity values equal to zero for the term *kill* with POS label n were obtained by adding polarity values assigned to each sense. Therefore, these obtained polarity values do not reflect the actual polarity of the term *kill*. We obtained the same positive and negative polarity values for *kill* when we add the polarities of all senses associated to this term with POS label v. The value 2.125 for both positive and negative polarity is wrong. Therefore, obtaining correct polarity values by adding polarity values associated with senses for each POS label is not guaranteed.

Hence, the values were normalized, polarity values were sought in the lists of positive and negative words, and some calculations were made in order to obtain the polarity values that best

Table 3. The preliminary polarity values of *kill* for each POS label

POS	Positive polarity values	Negative polarity values
n	0	0
v	2.125	2.125

Table 4. The final polarity values of *kill* for each POS label

POS	Positive polarity values	Negative polarity values
n	0	1
v	0.3375	1.2682

characterize the term *kill*. Thus, we obtained the polarity values of *kill* shown in Table 4. These are the polarity values included in SentiWordNet 4.0.

In the construction of SentiWordNet 4.0, we calculated the polarity of some terms with zero polarity in SentiWordNet 3.0. We used the steps outlined in Figure 2, taking into account the synonyms and antonyms of terms. For instance, the term *abductor* does not appear in the list of negative words despite being a negative term. Therefore, it was necessary to consider the polarities of its synonyms and antonyms. Thus, the calculated scores of the term *abductor* are 0.3016 for the positive polarity value and 0.8222 for the negative polarity value. This positive score surprised us, but it makes sense for such sentences as *He is the abductor of my heart*. Hence, the term *abductor* is related to other terms that despite being negative, express some positive polarity in some contexts.

5 Using the Developed Resources in Opinion Polarity Detection

In this section we show how the developed resources can improve the opinion polarity detection. To do this, we take as its starting point the application PosNeg Opinion and the general scheme on which it was developed [14]. First, we describe the general ideas of the scheme and the PosNeg Opinion application. After that we explain how to transform the scheme and application [14] using the developed resources.

The PosNeg Opinion application follows the general schema for unsupervised opinion polarity detection presented in Figure 4. This schema consists of five stages and determines the polarity of sentences taking into account the polarity of all meanings of each word in the sentence.

Stage 1 is responsible for reading the opinions, it selects the terms that provide useful information and eliminates stopwords. Each term is lemmatized and lexically disambiguated in Stage 2. If we process Spanish opinions, each Spanish term has to be translated to English in Stage 3 and its polarity values are obtained from the original SentiWordNet. The polarities of the terms are calculated considering the polarity values of all meanings of the terms in Stage 4. In Stage 5, the

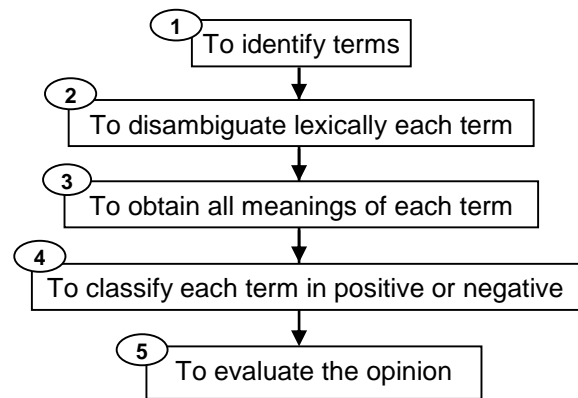


Fig. 4. General scheme for opinion polarity detection (Source: [14])

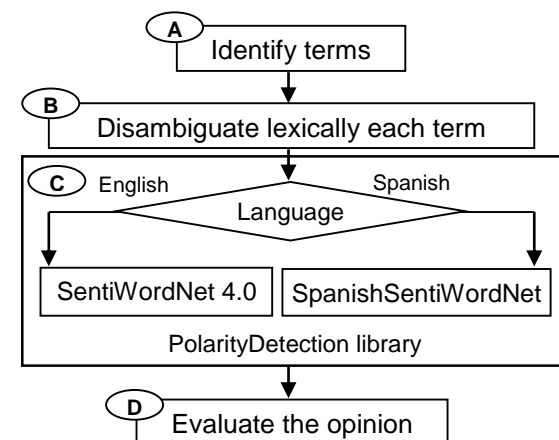


Fig. 5. General scheme for opinion polarity detection using PolarityDetection library

polarity of the opinion is calculated, considering the polarity of its terms.

To improve the efficiency and effectiveness of the PosNeg Opinion schema and application, we replace Stages 3 and 4 by using the developed resources: SentiWordNet 4.0, Intralinguistic Index, and SpanishSentiWordNet. To facilitate the application of these resources in this software or others with similar purposes, the PolarityDetection library was created. This library encloses the created resources.

PosNeg Opinion 2.0 was created considering the above modifications. PosNeg Opinion 2.0 only needs to transit through four stages. As we can see in Figure 5, Stages A, B, and D coincide with Stages 1, 2, and 5 from the original schema,

respectively. In Stage D, we use the PolarityDetection library for processing opinions in the English and Spanish languages using the SentiWordNet 4.0 and SpanishSentiWordNet.

The PolarityDetection library allows us to add positive and negative polarity values obtained from SentiWordNet 4.0 or SpanishSentiWordNet, respectively. Also, we use lists of modifier and negation words. Thus, it is possible to increase the polarity value of a term if the preceding term is a modifier word, or interchange the positive and negative polarity values if the term has a negative word precedent.

PosNeg Opinion 2.0 allows to process a lot of opinions in an unsupervised way and with a friendly interface. This application can be included in a more general opinion mining application as a module. It can easily communicate with other applications transferring information through XML files. It can also be easily extended to other languages, because it is only necessary to transfer the Intralinguistic Index from the specific language to the English language for obtaining the lexicon.

6 Study Case and Validation Results

In this section we consider a study case concerning 100 positive and 100 negative opinions manually labeled from a Yahoo discussion forum. This data set is applied to PosNeg Opinion 2.0 using the created resources presented in the previous sections.

Table 5 shows the distribution of the opinions by topics as well as the amount of positive and negative opinions. We chose different domains to prove that our proposal can work for any type of opinion. In addition, the selected opinions present

Table 5. The distribution of opinions by topics

Topics	Positive	Negative
Informatics products	14	20
Beauty products	10	5
Films and television	46	15
Social issues	15	45
Social networks	15	15

misspellings, writes in blocks, and strange characters which make them more real than other English corpora which mostly do not have these features.

On the one hand, the opinion polarity detection can be reduced to a classification problem where the classes are positive and negative; secondly, our opinion corpora is labeled. Hence, we can perform a supervised evaluation of PosNeg Opinion 2.0 using the classical measures for validating the performance of a classifier: Accuracy (ACC) and Recall or True Positive Rate (TPR), see Equation 1 and 2, respectively. TP means the number of true positives, FP means the number of false positives, FN means the number of false negatives, and TN means the number of true negatives:

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

The polarity of 200 opinions was detected using the PosNeg Opinion 2.0 application. The obtained Accuracy is 0.9599 and the obtained Recall is 0.9765, showing high quality results by applying the proposed scheme with the created resources.

7 Comparing with Related Works

In this section we compare our results with some related works mentioned in section 2. Only the related works with high Accuracy and Recall values were selected for the comparison.

In [6], the corpus used for evaluating the proposed system, named XRCE, contains 800 sentences belonging to the restaurant domain. The authors obtained a Recall value equal to 0.8625 when the XRCE was applied to the selected corpus for aspect term detection, and an Accuracy value equal to 0.7769 for aspect term polarity detection. Both measures yield values below 0.9.

We could not obtain the corpus used in [9]. Thus, in order to compare our results with the XRCE results, a corpus including 800 opinions about restaurants was created. The 800 opinions were taken from the TripAdvisor website [44]. The

using → 01158872-v (*utilize#1 utilize#1 use#1 employ#1 apply#1*)
loving → 01463965-a (*loving#1*)
nicer → 00984333-a (*squeamish#1 prissy#2 overnice#1 nice#4 dainty#4*)
windows → 04587648-n (*window#1*)
Looking → 02133435-v (*seem#1 look#2 appear#1*)
Forward → 00075442-r (*forward#3 ahead#2*)
using → 01158872-v (*utilize#1 utilize#1 use#1 employ#1 apply#1*)
latex → 15006118-n (*latex#1*)
editor → 10044879-n (*editor in chief#1 editor#1*)

Fig. 6. The synsets associated with the words present in the tweet

[01158872-v] (0.000 - 0.000) * 0.003819 → *use#1*...
 [01463965-a] (0.750 - 0.000) * 0.004263 → *loving#1*...
 [00984333-a] (0.000 - 0.375) * 0.004060 → *nice#4*...
 [04587648-n] (0.000 - 0.000) * 0.005603 → *window#1*...
 [02133435-v] (0.000 - 0.000) * 0.002388 → *look#2*...
 [00075442-r] (0.000 - 0.000) * 0.013473 → *forward#3*...
 [01158872-v] (0.000 - 0.000) * 0.003819 → *use#1*...
 [15006118-n] (0.000 - 0.000) * 0.007688 → *latex#1*...
 [10044879-n] (0.000 - 0.000) * 0.011033 → *editor#1*...

Fig. 7. Polarity score of each synset with its PageRank value

Table 6. Results of the application of PosNeg Opinion 2.0 to the corpus of restaurant opinions

Method	ACC	TPR
PosNeg Opinion 2.0	0.9367	0.9627

conditions for the comparisons are not the same, but both experiments include 800 opinions from the same domain.

High values of Accuracy and Recall are obtained when we apply PosNeg Opinion 2.0 to the created corpus, as we can see in Table 6. We only illustrate here that our approach has a better performance for a specific corpus from the same domain.

We will describe the example presented in [8] for showing disadvantages of sentiment polarity classification using a ranked WordNet and how it is possible to solve them using our approach. The starting point of the exemplification is the following tweet: “*Using Linux and loving it - so much nicer than windows... Looking forward to using the wysiwyg latex editor! :)*”. After cleaning the tweet,

the text to apply in the following step was: “*Using Linux and loving it so much nicer than windows Looking forward to using the wysiwyg latex editor*”.

Before making up the PPV and getting the polarity scores from SentiWordNet, a disambiguation process was applied with the aim of discovering the correct synset of each term. The synsets associated with the words present in the tweet are shown in Figure 6. After the disambiguation of the terms of the tweet, the following step is the building of PPV. The final step is weighting the SentiWordNet polarity score of each synset with its PageRank value, as it is shown in Figure 7. The polarity score is obtained as a result of the difference between the positive and negative scores. The final polarity score is 0.000186, i.e. the system assigns the positive class.

We directly use the example given in [8] to highlight the disadvantages of SentiWordNet and its influence on calculation of the polarity of opinions. As we can see in this example, if the user had omitted the word *loving*, the tweet would have been classified as negative, because the only terms that do not add zero polarity are *loving* and *nice*. We note also that *nice* is incorrectly classified, because the applied SentiWordNet assigned a negative polarity to *nice*. Errors like these in our approach are avoided by the introduced modifications in the SentiWordNet. Therefore, applications that depend on this resource will get better results if they use our SentiWordNet 4.0.

Also, we want to compare our approach to the unsupervised sentiment orientation system proposed in [9]. The experiments have been performed on both systems using 50 sentences of phone reviews which were collected from Amazon [45]. PosNeg Opinion 2.0 yielded better values of Accuracy and Recall than the system proposed in [9] as shown in Table 7.

Finally, we want to compare the opinion polarity detection results obtained over a Spanish corpus using our resources with those obtained using the resource eSOL proposed in [9]. The MuchoCine corpus [46] was the selected Spanish corpus for the experimentation. The reviews are written by web users instead of professional film critics. This increases the difficulty of the task because the sentences found in the documents may not always be grammatically correct, or they may include

Table 7. Results of the application of the system proposed in [9] and PosNeg Opinion 2.0 to the corpus of phone reviews

Method	ACC	TPR
System proposed in [9]	0.7401	0.7878
PosNeg Opinion 2.0	0.9414	0.9851

Table 8. Rating distribution

Rating	Number of reviews
1	351
2	923
3	1253
4	890
5	471
Total	3865

Table 9. Binary classification of the MuchoCine corpus

Classes	Number of reviews
Positive	1274
Negative	1361
Total	2635

Table 10. Obtained results applying eSQL and our proposed resources

Method	ACC	TPR
Using eSQL resource	0.6316	0.6393
Using our proposed resources	0.8860	0.9057

spelling mistakes or informal expressions. The corpus contains about two million words and an average of 546 words per review. The opinions are rated on a scale from 1 to 5 as shown in Table 8, where 1 means that the movie is very bad and 5 means that it is very good.

In our experiments we discarded the neutral examples. In this way, opinions rated with 3 were not considered, the opinions with ratings of 1 or 2 were considered as positive and those with ratings of 4 or 5 were considered as negative. Table 9 shows the class distribution of the binary classification of the reviews.

Table 10 shows the validation results applying eSQL and our proposed resources to the MuchoCine corpus. As seen in Table 10, our proposal significantly exceeds the ACC and TPR values obtained using the eSQL resource.

8 Conclusions

Two new lexical resources, SentiWordNet 4.0 and SpanishSentiWordNet, have been proposed in this paper.

The changes in the format of the SentiWordNet 3.0, the five stages and their respective algorithms, defined for reallocating the scores associated to the meanings of the terms, allowed to obtain the new SentiWordNet 4.0 that has 5495 matches with the list of words reported in [33], thus ensuring greater certainty in the assigned positivity and negativity values of the terms.

The Intralinguistic Index format was transformed by removing the field that stored the identifiers of the different semantic relations of the terms and the repeated meanings of the English words thus achieving the eradication of Intralinguistic Index disadvantages for opinion mining applications.

The SpanishSentiWordNet was created exploiting the advantages of Intralinguistic Index and SentiWordNet 4.0. It includes 43525 terms; of these, 30289 nouns, 8664 adjectives, and 4572 verbs.

The integration of the proposed resources was solved by combining them in the new PolarityDetection library which was integrated to PosNeg Opinion 2.0 and facilitates, in general, the development of polarity detection applications in both English and Spanish languages.

The experimental results showed that adding the developed resources to the PosNeg Opinion application can improve the polarity detection process, both in the English and Spanish languages, indicating the effectiveness of the developed resources. In the experimental study we showed that our approach overcame the ACC and TPR values with respect to the values obtained from the related works. The ACC and TRP values were always bigger than 0.9, except the ACC value equal to 0.8860 obtained by classifying the opinions of the MuchoCine corpus.

References

1. **Pang, B. & Lee, L. (2008).** Opinion mining and sentiment analysis. *Found. Trends Inf Retr.* Vol. 2, No. 1-2, pp. 1–135. DOI: 10.1561/1500000011.
2. **Liu, B. (2010).** *Handbook of Natural Language Processing: Sentiment analysis and subjectivity.* 2 ed, Taylor & Francis Group.
3. **Liu, B. (2012).** *Sentiment Analysis and Opinion Mining.* Morgan & Claypool Publishers.
4. **Belbachir, F. & Missen, M.M.S. (2013).** Effectiveness gain of polarity detection through topic domains. *Proc. of the 7th International Conference on Advances in Semantic Processing*, pp. 62–67.
5. **Macdonald, C., Ounis, I., & Soboroff, I. (2007).** Overview of the TREC Blog Track. *Proc. of TREC*, 7, pp. 31–43.
6. **Molina-González, M., Martínez-Cámara, E., Martín-Valdivia, M., & Perea-Ortega, J. (2013).** Semantic orientation for polarity classification in Spanish reviews. *Expert Systems with Applications*, Vol. 40, No. 18, pp. 7250–7257. DOI: 10.1016/j.eswa.2013.06.076.
7. **Sharma, R., Nigam, S., & Jain, R. (2014).** Polarity detection at sentence level. *International Journal of Computer Applications*, Vol. 86, No. 11.
8. **Montejo-Ráez, A., Martínez-Cámara, E., Martín-Valdivia, M., & Ureña-López, L. (2014).** Ranked WordNet graph for sentiment polarity classification in Twitter. *Computer Speech & Language*, Vol. 28, No. 1, pp. 93–107.
9. **Brun, C., Nicoleta, D., & Roux, C. (2014).** XRCE: Hybrid Classification for Aspect-based Sentiment Analysis. *Proc. of the SemEval*.
10. **Anta, A., Núñez, L., Morere, P., & Santos, A. (2013).** Sentiment analysis and topic detection of Spanish tweets: A comparative study of NLP techniques. *Procesamiento del lenguaje natural*, Vol. 50, pp. 45–52.
11. **Gul, N. (2014).** SCTUR: A Sentiment Classification Technique for URDU Text. *International Journal of Computer and Communication System Engineering (IJCCSE)*, Vol. 1, No. 3, pp. 97–101.
12. **Tang, H., Tan, S., & Cheng, X. (2009).** A survey on sentiment detection of reviews. *Expert Systems with Applications*, Vol. 36, No. 7, pp. 10760–10773. DOI: 10.1016/j.eswa.2009.02.063.
13. **Paltoglou, G. & Thelwall, M. (2012).** Twitter, MySpace, Digg: Unsupervised sentiment analysis in social media. *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 3, No. 4. DOI: 10.1145/2337542.2337551.
14. **Amores, M., Arco, L., & Artilles, N. (2015).** PosNeg opinion: Una herramienta para gestionar comentarios de la Web. *Revista Cubana de Ciencias Informáticas*, Vol. 9, No. 1, pp. 20–31.
15. **Esuli, A. & Sebastiani, F. (2006).** Sentiwordnet: A publicly available lexical resource for opinion mining. *Proc. of LREC*, 6, pp. 417–422.
16. **Strapparava, C. & Valitutti, A. (2004).** WordNet Affect: an Affective Extension of WordNet. *Proc. of LREC*, 4, pp. 1083–1086.
17. **Stone, P.J., Dunphy, D.C., & Smith, M.S. (1966).** *The General Inquirer: A Computer Approach to Content Analysis.*
18. **Yang, K., Yu, N., & Zhang, H. (2007).** WIDIT in TREC 2007 Blog Track: Combining Lexicon-Based Methods to Detect Opinionated Blogs. *Proc. of TREC*.
19. **Zhou, G., Joshi, H., & Bayrak, C. (2007).** Topic Categorization for Relevancy and Opinion Detection. *Proc. of TREC*.
20. **Brooke, J., Tofiloski, M., & Taboada, M. (2009).** Cross-Linguistic Sentiment Analysis: From English to Spanish. *Proc. of RANLP*.
21. **Martinez-Camara, E., Martin-Valdivia, M., & Ureña-Lopez, L. (2011).** Opinion classification techniques applied to a Spanish corpus. *Proc. of NLDB*, pp. 169–176.
22. **Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásques, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., & Gordon, J. (2012).** Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets. *Proc. of MICAI'12*, pp. 1–14. DOI: 10.1007/978-3-642-37807-2_1.
23. **Xie, L., Zhou, M., & Sun, M. (2012).** Hierarchical structure based hybrid approach to sentiment analysis of Chinese micro blog and its feature extraction. *Journal of Chinese Information Processing*, Vol. 26, No. 1, pp. 73–82.
24. **Ishijima, H., Kazumi, T., & Maeda, A. (2014).** Sentiment Analysis for the Japanese Stock Market, *Forthcoming in Global Business and Economics Review*, Vol. 17, No. 3. DOI: 10.1504/GBER.2015.070303.
25. **Abdul-Mageed, M., Kübler, S., & Diab, M. (2012).** Samar: A system for subjectivity and sentiment analysis of Arabic social media. *Proc. of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pp. 19–28.
26. **Erdmann, M., Ikeda, K., Ishizaki, H., Hattori, G., & Takishima, Y. (2014).** Feature Based Sentiment Analysis of Tweets in Multiple Languages. *Proc. of the Web Information Systems Engineering*

- Conference (WISE), pp. 109–124. DOI: 10.1007/978-3-319-11746-1_8.
27. **Balahur, A., Turchi, M., Steinberger, R., Perea-Ortega, J.M., Jacquet, G., Küçük, D., Zavarella, V., & El Ghali, A. (2014).** Resource Creation and Evaluation for Multilingual Sentiment Analysis in Social Media Texts. *Proc. of the 9th edition of the Language Resources and Evaluation Conference (LREC)*.
 28. **Schmid, H. (1994).** *Treetagger*. TC project at the Institute for Computational Linguistics of the University of Stuttgart.
 29. **Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990).** Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, Vol. 3, No. 4, pp. 235–244. DOI: 10.1093/ijl/3.4.235.
 30. **Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. (2014).** Semeval-2014 task 4: Aspect based sentiment analysis. *Proc. of the International Workshop on Semantic Evaluation (SemEval)*.
 31. **Ait-Mokhtar, S., Chanod, J., & Roux, C. (2001).** A multi-input dependency parser. *Proc. of IWPT*.
 32. **Brun, C. (2011).** Detecting opinions using deep syntactic analysis. *Proc. of RANLP*, pp. 392–398.
 33. **Hu, M. & Liu, B. (2004).** Mining and summarizing customer reviews. *Proc. of the tenth SIGKDD*, ACM.
 34. **Manning, D.C. (2011).** Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics? *Proc. of the 12th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*.
 35. **K. Gimpel, Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., & Smith, N.A. (2011).** Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments. *Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT)*, Vol. 2.
 36. **Hu, M. & Liu, B. (2004).** Mining and summarizing customer reviews. *Proc. of the tenth International Conference on Knowledge Discovery and Data Mining (KDD)*, ACM, New York, USA, pp. 168–177.
 37. **Montejo-Ráez, A., Díaz-Galiano, M.C., Perea-Ortega, J.M., & Ureña-López, L.A. (2013).** Spanish knowledge base generation for polarity classification from masses. *Proc. of the 22nd International Conference on World Wide Web Companion*, pp. 571–578. DOI: 10.1145/2487788.2487996.
 38. **Cambria, E., Havasi, C., & Hussain, A. (2012).** SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis. *Proc. of the FLAIRS Conference*, pp. 202–207.
 39. **Esuli, A. & Sebastiani, F. (2006).** SENTIWORDNET: A high-coverage lexical resource for opinion mining. *Proc. of the 5th Conference on Language Resources and Evaluation (LREC)*, pp. 1–26.
 40. **Esuli, A. (2008).** *Automatic Generation of Lexical Resources for Opinion Mining: Models, Algorithms, and Applications*. Ph.D. thesis, Scuola di Dottorato in Ingegneria “Leonardo da Vinci” University of Pisa, Pisa, IT.
 41. **Baccianella, S., Esuli, A., & Sebastiani, F. (2010).** SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *Proc. of LREC*, 10, pp. 2200–2204.
 42. **Vossen, P. (2004).** EuroWordnet: a multilingual database of autonomous and language-specific Wordnets connected via an Inter-LinguallIndex. *International Journal of Lexicography*, Vol. 17, No. 2, pp. 161–173. DOI: 10.1093/ijl/17.2.161.
 43. **Vossen, P. (2011),** *EuroWordnet*. Technical Report, University of Amsterdam.
 44. <http://www.tripadvisor.com>
 45. <http://www.amazon.com>
 46. <http://www.csie.ntu.edu.tw/~cjlin>
- Mario Amores** received his B.Sc. and M.Sc. in Computer Science from the Central University of Las Villas (UCLV), Cuba, in 2013 and 2016, respectively. He is currently a professor at the Computer Sciences Department of the Central University of Las Villas, Cuba, and works as researcher at the Artificial Intelligence Lab of the Center for Informatics Studies in that University. He is interested in text and opinion mining, machine learning, and artificial intelligence techniques.
- Leticia Arco** received her B.Sc. and M.Sc. in Computer Science from the Central University of Las Villas (UCLV), Cuba, in 2001 and 2005, respectively. In 2009 she obtained her Ph.D. degree in technical sciences from the National Committee of Automatics and Computing and the Central University of Las Villas, Cuba, in collaboration with the University of Oldenburg, Germany. She fulfilled a postdoctoral stay at the

Vrije Universiteit Brussel, Belgium, in 2015. She was the vice-dean of research and graduated courses at the Faculty of Mathematics, Physics, and Computer Sciences (UCLV) from 2008 to 2014. She is currently a professor at the Computer Sciences Department of the Central University of Las Villas, Cuba. She is interested in text and opinion mining, machine learning, and artificial intelligence techniques.

Claudia Borroto received B.Sc. in Computer Science from the Central University of Las Villas (UCLV), Cuba, in 2014. She is currently a teacher at the Department of Informatics of the University of Ciego de Ávila “Máximo Gómez Báez” (UNICA), Cuba. She is interested in text and opinion mining, machine learning and artificial intelligence techniques.

*Article received on 13/11/2015; accepted on 08/04/2016.
Corresponding author is Mario Amores.*