AspectSA: Unsupervised system for aspect based sentiment analysis in Spanish

Sistema no supervisado para el análisis de sentimiento basado en aspectos en español

Carlos Henríquez Miranda[1], Edgardo Buelvas [2]

PhD. Profesor de Tiempo Completo, Universidad Autónoma del Caribe, Grupo de Investigación SINT-UAC,
[2]PhD. Profesor de Tiempo Completo, Universidad Autónoma del Caribe, Grupo de Investigación SINT-UAC,
Barranquilla. E-mail: carlos.henriquez@uac.edu.co

Recibido: 19/01/2019
Aceptado: 01/04/2019

Cite this article as: *: C. Henríquez Miranda, E. Buelvas "AspectSA: Unsupervised system for aspect based sentiment analysis in Spanish", Prospectiva, Vol 17, N° 1, 87-95, 2019.*

ABSTRACT

This paper describes an unsupervised system for sentiment analysis is presented in Spanish. The system performs a complete fine grain analysis, where the most important characteristics or aspects of an opinion are identified in order to determine their sentiment associated. The unsupervised approach used allows to extract, identify and sentiment classify, from the analysis of opinions in Spanish in a specific domain, allowing to scale to another language and domain with great ease. For the validation of AspectSA, several experiments were carried out using corpus of opinions in the restaurant domain. The results obtained exceeded the majority of existing systems for the Spanish language.

Key words: *Sentiment Analysis; Unsupervised; Aspect Based; Opinion Mining; NLP*

RESUMEN

En este artículo se presenta un sistema no supervisado para el análisis de sentimientos en español. El sistema realiza un análisis completo de grano fino, en donde se identifican las características o aspectos más importantes de una opinión para poder determinar su sentimiento asociado. El enfoque no supervisado utilizado permite extraer, identificar y clasificar sentimientos, a partir del análisis de opiniones en español en un domino especifico, permitiendo escalar a otro lenguaje y dominio con gran facilidad. Para la validación de AspectSA se realizaron varios experimentos utilizando corpus de opiniones en los dominios de restaurante. Los resultados obtenidos superaron a la mayoría de sistemas existentes para el lenguaje español.

Palabras clave: *Análisis de Sentimientos; No supervisado; Basado en aspectos; Minería de opiniones;PLN*

http://dx.doi.org/10.15665/rp.v17i1.1961

1. INTRODUCTION

Today, there is a large amount of data, so large and complex, that traditional data processing tools have become obsolete to deal with them properly. For example, the Internet produces an excessive volume of data, due to the massive use of social networks, messaging services, wikis, blogs, and e-commerce, among others. All these data are attractive for organizations of all kinds, but the extraction and its respective processing in a traditional way, is complex and difficult to perform [1][2]. In order to analyze and manage this data produced by people, there are great work fronts to find models, techniques and tools that allow the analysis of texts automatically. Recent research has led to a specific area of natural language processing (NLP), known as sentiment analysis (SA) or opinion mining (OM). The SA seeks to analyze the opinions, sentiments, values, attitudes and emotions of people towards entities such as products, services, organizations, individuals, problems, events, themes and their attributes [3].

The AS is to analyze a text, usually an opinion or comment, and determine in that text a positive or negative assessment, which is known as sentiment classification or polarity assignment [4]. To analyze the text there are basically three approaches: The document-level approach that consists of sentiment classification of a large text [5], the phrase-level approach that does it only in one prayer and the aspect level approach, which performs a fine-grained analysis on a whole document, identifying the most important characteristics or aspects of the text and assigning sentiment [6].

Currently, the vast majority of approaches for SA detect sentiments at a general level in a complete document [7]. However, these are inconclusive for those involved since they need to know in detail what is said about their product or service [8]. For the above, interest in Aspect-Based Sentiment Analysis (ABSA) has grown in recent years [9], being more active in the English language than in other languages such as Spanish, Chinese and Czech among others [10] [11].

The ABSA aims to identify the properties (aspects) of an entity and determine the polarity of that entity. An aspect is an attribute or component of an entity. For example, in the phrase, "*The beef of this restaurant is exquisite*", the aspect is "*beef*", the entity is "*restaurant*" and the associated sentiment is "*exquisite*" that has "*positive*" polarity. Within this approach two types of aspects are distinguished, the first one refers to the explicit aspects that are words in the document that directly denote the objective of the opinion. The second, is the implicit aspect, this represents the objective of opinion of a document, but that does not appear explicitly in the text [12]. The vast majority of contributions of the literature in ABSA concentrate on identifying only the explicit aspects to come from supervised and dictionary-based approaches [13], leaving in the background the identification, extraction and sentiment classification of the implicit aspects [14].

In this article, we show the details of the construction of *AspectSA*, an unsupervised system for the analysis of sentiment based on aspects in Spanish, that allows to automatically extract the aspects (explicit and implicit) of a text. *AspectSA* is based on a semantic model that integrates PLN, ontologies, semantic similarity and unsupervised automatic learning, which seeks to minimize human participation throughout the process. This system is composed of four (4) components: The language processor that performs preprocessing tasks of the text, the aspects extractor that identifies and extracts the explicit and implicit aspects, the sentiment identifier that determines the characteristic that expresses the sentimental situation and the sentiment classifier that determines the polarity.

The rest of the article is organized as follows. Section II deals with background and similar work; Section III describes the system architecture; Section IV shows the implementation of the system and its validation. Finally, the conclusions in the section in section V.

2. BACKGROUND AND RELATED WORKS.

The ABSA or also known as feature-based sentiment analysis [15], aims to identify the properties or characteristics of an entity and determine the expressed polarity of each aspect of that entity. According to [6] there are usually two tasks related to ABSA. The first task is related to detecting and extracting aspects of an entity in a given text and the second determining the sentiment associated with that aspect.

For the extraction of aspects there are different approaches treated in the literature. Those that use a predetermined list of aspects[16] ,those count of names to calculate their frequency [17] and those that take advantage of the relationship between sentiment and aspects [18]. In addition, there are more advanced approaches such as supervised learning [19] and models based on probabilistic inference [20].Of all the previous approaches, the vast majority does not take into account the meaning of the aspects, that is, their extraction and identification is made taking into account the words and not the concept they represent [21]. To deal with this situation, in the proposed system (*AspectSA*), semantic techniques based on ontologies and semantic similarity are used to extract the important aspects in the text taking into account their meaning.

With respect to the other task, to determine the sentiment associated, basically two strategies are used. The one based on machine learning and the lexical-based one [22]. The machine learning approach is based on the application of an algorithm that learns from a set of example data and the lexical-based strategy, it needs a sentiments lexicon or dictionaries of words with their polarity to be able to process them [23]. Within the machine learning approach is supervised and unsupervised learning; the first depends on the existence of previously labeled training documents, and the second does not require prior knowledge of labeled data. Supervised learning prevails regarding unsupervised and tends to achieve better classification results, due to a large number of documents labeled training. However, sometimes it is difficult to have these documents labeled because a human should normally be used for this task [13]. In the strategy based on the lexicon, although there are many dictionaries composed mainly of adjectives, these are very general for an AS at the level of the aspects, since they do not take into account neither the environment nor the domain that surrounds the aspect.

In the literature, we found few references to ABSA in Spanish, much less on implicit aspects [24]. Of the works found, most are limited to applying the same techniques used in the English language [11]. Within the literature related to ABSA, there are works such as: [25] carries out an analysis supported by ontologies in the cinema and hotels domain in Portuguese, [7] presents a supervised approach in restaurant reviews in Czech, [26] proposes a system in the English language based on the GINI index on cinema, [27] which is based on training all aspects of a particular sentence and capitalizing on inter-aspect dependency modeling. Recently unsupervised approaches have been used such as [28] and [29] in English and [30] in multilingual. Another semi-supervised approach and special consideration are given in [6], a system in Spanish for ABSA that combines an unsupervised model for extraction of aspects and supervised machine learning for classification of sentiment.

Our work is different from all these works is the proposed system is based on automatic unsupervised learning does not depend on the data labeled or dictionaries polarity, so you try to minimize human involvement at all the process.

3. SYSTEM ARCHITECTURE

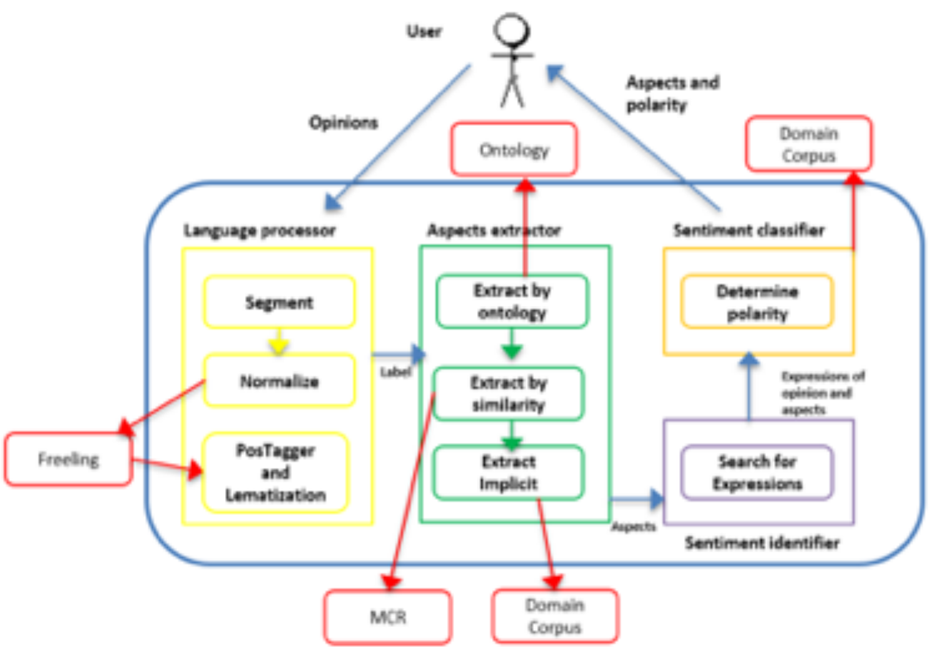
In Figure 1, the architecture of the system is shown. The architecture shows four (4) components: The language processor, Aspects extractor, the Sentiment identifier and Sentiment classifier.

In general terms, the system starts its operation when the user enters a set of opinions the domain particular and language. With the set of opinions entered, the language processing component initially performs a segmentation consisting of identifying the number of incoming sentences. The result is sent to the next stage that is responsible for normalizing the text. What is done here is to lower all the words, delete all the symbols and finally verify that each sentence within the set of sentences ends at a point.

The next step is to send each of the sentences to the external *Freeling* [31] component that is responsible for making the grammar labeling process (PosTagger) and lematization. This processed data is sent to the aspects extractor component.

The aspects extractor component is responsible for the extraction of the explicit and implicit aspects of the opinions. The above is done in three stages. The first stage looks for explicit aspects in each sentence through an ontology of domain taking those words with "noun" category and looking for their respective lexical coincidence between the classes and individuals of the ontology. The next stage seeks more aspects explicit from those words "nouns" that were not found in the ontology, for this a process of semantic similarity is made from queries to a lexical database based on *WordNet* MCR[32]. In the last stage implicit aspects are looked for in sentences where there is no explicit aspect. For this, a co-

Figure 1. System Architecture
Figura 1. Arquitectura del sistema



Source: Authors.

occurrence matrix is constructed between possible explicit aspects and nominal expressions based on a double propagation process and access to a domain corpus.

With the output of the previous component, the sentiment identifier component takes the list of explicit aspects and looks for the expressions of opinion close to it with a sliding window of length two (defined experimentally for the restaurant domain). The expressions of opinion will be the words with grammatical category adverb and adjective close to the aspect. Here it is possible that the aspect does not have an associated expression, either because it is outside the window or does not appear within the opinion. In this case they are marked without expression of opinion and we are processed by the following component.

Finally, the sentiment classification component obtains a list of explicit aspects with their respective expressions of opinion and a list of implicit aspects. In this component a list of sentiment seeds *("excelente", "bueno", "malo", "pésimo", "indiferente")* is defined so that along with the appearance and the expressions of opinion the polarity of the aspect is calculated using the measure of association Pointwise Mutual Information (PMI) [33] .The PMI is a measure of association that is obtained between two words, *x* and *y*, by the probability that the two words appear together divided by the probabilities of each word in individual form (see Formula 1).

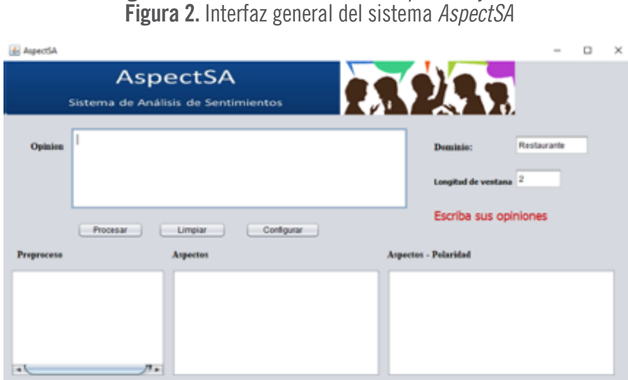
$$PMI(x,y) = \log 2\left(\frac{P(x,y)}{P(x)P(y)}\right) \quad (1)$$

Source: [33]

RESULTS AND VALIDATION

AspectSA is a software built under the JAVA programming language that uses components from two external tools (see Figure 2). The first tool is the *Freeling 4.0* library that allows the realization of grammatical labeling and lematization. The second tool is the multilingual lexical knowledge base (MCR) of wide coverage based on *Wordnet* in Spanish. The latter is used for the process of semantic similarity that allows to extract aspects.

Figure 2. General interface of the *AspectSA* system
Figura 2. Interfaz general del sistema *AspectSA*



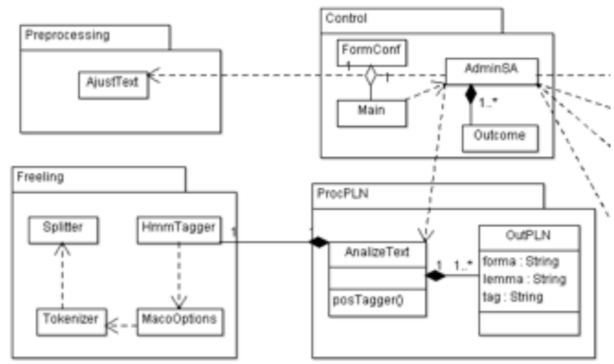
Source: Authors.

In the following sections, a description will be given of the main components that make up each *AspectSA* module and will also describe how the architecture of the modules interacts with external systems to carry out the final process.

4.1 Language processor module

This module is in charge of analyzing the input text and producing a list of word tags and lematized words. Figure 3 shows the package diagram that includes the relationships between the classes of this component.

Figure 3. Package diagram of the Language processor module
Figura 3. Diagrama de paquetes del módulo Procesador de lenguaje



Source: Authors. Other packages

In Figure 3 it can be seen that four packages are involved in this module: Control, Preprocessing, ProcPLN and Freeling. The Control package manages the AdminSA, Main, FormConf and Outcome classes. AdminSA is the main controller class that receives the input data that has been captured by the Main class from an independent graphical interface. This class (AdminSA) will be present throughout the process and will save the final output through objects of the Outcome class. These initial data meet all the requirements presented in the Architecture. The Main class receives the text and validates if there is at least one written opinion. For the system, each opinion ends at one point. After verifying if there is an opinion, the Main class sends the written text to an object of the AdminSA class. It should be noted that the system receives the texts even if they are poorly written, which can cause errors in the acquisition of the aspects and polarity later.

In *AdminSA*, an object of the *AJustText* class is used to perform the segmentation (divide all the text into sentences and sentences into words) and standardization (which consists of basically lowercase, erase all the symbols and place the symbol for each sentence point).

Then an object of the *AnalizeText* class is used, which handles all the remaining process of this component. This object invokes the *posTagger()* method where the connection to the *Freeling* package is made. In the *posTagger()* method, an object of the *HmmTager* class of the *Freeling* package is used to perform the labeling and lematization process. *Freeling* is an open-source library, external to the system, for automatic multilingual processing, it provides a range of linguistic analysis services for various languages. In this paper, *Freeling* libraries were selected in its version 4.0 due to its great robustness in this type of processes associated with the Spanish language.

The *HmmTagger* object abstracts a tagger based on a hidden *Markov* model that assigns the most likely tag for the word, based on assigning tags for a sentence as a whole, instead of looking for a tag for each word separately. For the assignment of the label, *Freeling* is based on a file that contains the statistical data for the *Markov* model, plus some additional data to smooth the missing values, initial probabilities, transition probabilities, lexical probabilities, among others. With respect to the *FormConf* class, this allows you to change the initial configuration of the system that is with default values for the restaurant domain and window length equal to two (2).

4.2 Aspects extractor module

In this module the aspects are identified and extracted from the entry of a list of opinions that are stored in objects of the *OutPLN* class (words, lemmas and their labels). Figure 4 shows the package diagram that includes the relationships between the classes of this component.

From the *AdminSA* class, an object of the *FindAspect* class is used, which invokes the *getDataOntology* method. This method obtains a list of concepts (ontology classes) each related to a list of individuals of the ontology. The set of classes and their respective individuals are stored in a list of objects of the *OntologyClass* class.

Subsequently, with the *FindAspect* type object, the *extractAspect* method is invoked, which receives as parameters (information) the *OntologyClass* list and the *OutPLN* list. The result of this process is a set of explicit aspects according to the defined architecture that is stored in a list of objects of the *OutCome* class. An object of the *OutCome* class will store every aspect along with the expressions of opinion and their polarity.

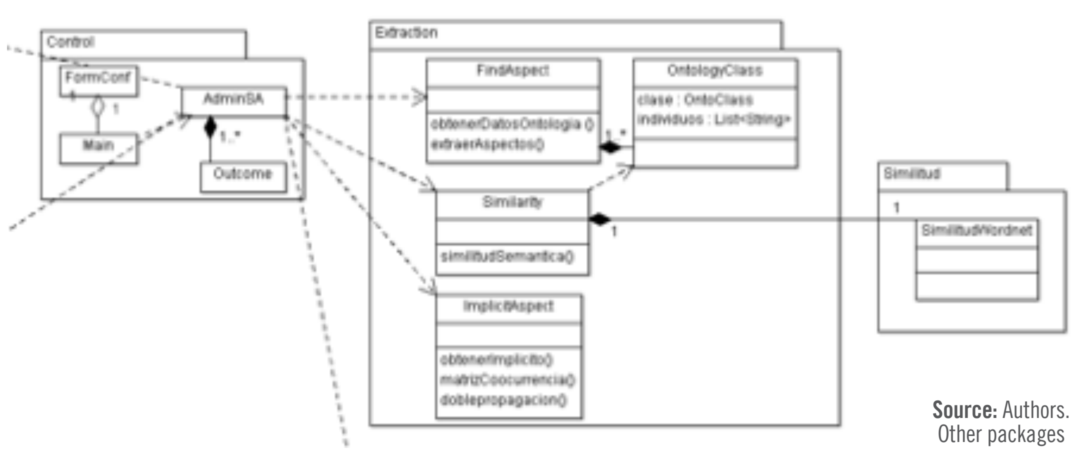
After the previous process the system now looks for more aspects applying the process of semantic similarity. For this process the multilingual lexical knowledge base (MCR) of wide coverage based on *Wordnet* was used. MCR integrates six different versions of the English *Wordnet* (from 1.6 to 3.0) and also *Wordnet* in Spanish, Catalan, and Italian, together with more than one million semantic relations between concepts as well as semantic properties of different ontologies. It should be noted that for the proposed system the database is downloaded with all the necessary information to access the MCR.

For this process the collection of objects of the *OutPLN* class is taken and those expressions with a substantive label of each opinion are obtained that were not cataloged as aspects in the previous process. That is, they were not found in the ontology. Then an object of the *Similarity* class is used to invoke the s*emanticSimilarity* method that returns a list of words considered aspects when the similarity calculation is higher than a threshold. For the calculation of semantic similarity based on MCR, the Wu & palmer measurement was used, which is suitable for working with the selected taxonomy and depends on the length and depth of the concepts.

The access to MCR is done through an object of the class *SimlitudWordnet* that invokes the *searchSynset* method to find the set of synonyms (Synset) of a word and the *obtainHiperonimos* method that obtains all the most general concepts of a Synset (the superior concepts in a taxonomic relation is_un). The Synset found of a concept come in a particular domain (gastronomy, food, construction, among others) according to the structure of *Wordnet*. With the Synset and Hyperonyms found, the distance between each concept is calculated and there is the common hyperonym concept that allows to calculate the semantic similarity. Figure 5 shows the query to the database in the searchSynset method and in Figure 6 the query made by the *obtainHiperonimos* method is shown.

The last process performed in this module is to find out if there are implicit aspects. The implicit aspects will be obtained from those opinions where no explicit aspect has been identified in the previous processes.

Figure 4. Package diagram of the Aspects extractor module
Figura 4. Diagrama de paquetes del módulo Extractor de aspectos



Source: Authors.
Other packages

Figure 5. Query for Synset in MCR
Figura 5. Consulta para obtener los Synset en MCR

```
String sql = "select syn.offset, \n" +
             "    syn.pos, \n" +
             "       ili_dom.domain \n" +
             "  from synsets syn \n" +
             "  join variantes var \n" +
             "    on var.offset = syn.offset \n" +
             "   and var.pos = syn.pos \n" +
          "left join synsets_ili syn_ili \n" +
             "    on syn_ili.offset = syn.offset \n" +
             "   and syn_ili.pos = syn.pos \n" +
          "left join ili_dominios ili_dom \n" +
             "    on ili_dom.ilioffset = syn_ili.ilioffset \n" +
             "   and ili_dom.ilipos = syn_ili.pos \n" +
             "   and ili_dom.iliwnid = syn_ili.iliwnid \n" +
             "where var.word = ? \n" +
             "   and var.pos = ? ";
```

Source: Authors.

To find the implicit aspects, it is necessary to invoke two processes outside the *AspectSA* system. The first has as input a list of objects of classes of the ontology and uses the technique of double propagation to find the nominal expressions whose label is adjective, adverb and verb that are related to the aspect. To know if they are related take the nominal expressions that appear in an opinion accompanying an aspect. This technique of double propagation is applied in a corpus of opinion of more than 50000 opinions where it starts first with the aspects to find the nominal expressions. When the process ends, it begins with the nominal expressions found and seeks a relationship with words that are substantive in the opinion. The process ends when there is no new aspect or new expression. With the list of aspects and expressions, a co-occurrence matrix is constructed, where the aspect appears along with the nominal expression in each opinion of the corpus. At the end we have a matrix of aspects against nominal expressions where we have the number of times they appear together in the corpus. In Figure 7 we can see an example of the result of the process where the "*comida*" aspect appears in the restaurant domain separated with the nominal expressions by the symbol "#" and the number of co-occurrences between the appearance and the nominal expression. For example, the "*comida*" aspect appears with the expression "*inmejorable*" 150 times in the corpus.

Figure 6. Query for the Hyperonyms in MCR
Figura 6. Consulta para obtener los Hiperónimos en MCR

```
String sql = "select syn_origen.offset, \n" +
             "      syn_origen.pos, \n" +
             "       ili_dom_origen.domain \n" +
             "       \n" +
             "  from synsets syn_destino \n" +
             "  join relaciones_detalles rd \n" +
             "    on rd.targetSynset = syn_destino.offset \n" +
             "   and rd.targetpos = syn_destino.pos \n" +
             "  join relaciones r \n" +
             "    on r.id = rd.relation \n" +
             "  join grupos_relaciones gr \n" +
             "    on gr.id = r.grup \n" +
             "  join synsets syn_origen \n" +
             "    on syn_origen.offset = rd.sourceSynset \n" +
             "   and syn_origen.pos = rd.sourcepos \n" +
          "left join synsets_ili syn_ili_origen \n" +
             "    on syn_ili_origen.offset = syn_origen.offset \n" +
             "   and syn_ili_origen.pos = syn_origen.pos \n" +
          "left join ili_dominios ili_dom_origen \n" +
             "    on ili_dom_origen.ilioffset = syn_ili_origen.ilioffset \n" +
             "   and ili_dom_origen.ilipos = syn_ili_origen.pos \n" +
             "   and ili_dom_origen.iliwnid = syn_ili_origen.iliwnid \n" +
             " where syn_destino.offset = ? \n" +
             "   and syn_destino.pos = ? \n" +
             "   and gr.name = 'Hyponyms'";
```

Source: Authors.

With the result of the two processes are now the implicit aspects of an opinion using the *ImplicitAspect* class and invoking the *obtainImplicito* method.

Figure 7. Example of the co-occurrence matrix for the *"comida"* aspect
Figura 7. Ejemplo de la matriz de coocurrencias para el aspecto *"comida"*

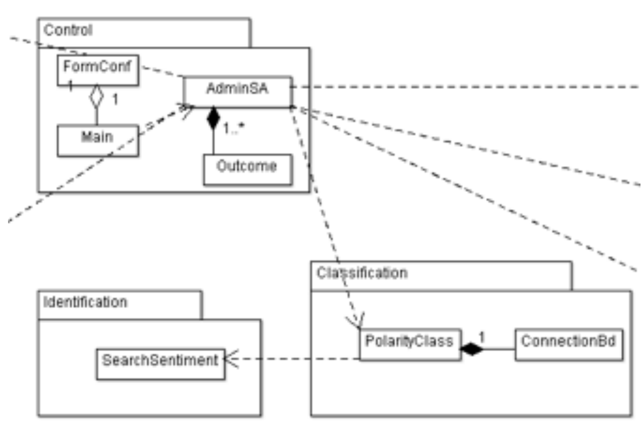
```
comida#tambien#428
comida#bodegó#78
comida#inmejorable#150
comida#serio#36
comida#antiguo#31
comida#uedo#264
comida#abundante#1218
comida#temprano#104
comida#ueda#738
comida#presionar#11
comida#bodego#58
comida#pretencioso#45
comida#provoleta#83
comida#vegetariano#45
comida#cortés#15
comida#festejar#143
comida#honestamente#14
comida#fumar#20
comida#comid#18768
comida#exicano#97
comida#excelentemente#26
comida#logrado#78
comida#seguramente#65
```

Source: Authors.

4.3 Sentiment identifier module

In this module are the expressions of opinion that are related to the aspects found in the previous module. Figure 8 shows the package diagram that includes the relationships between the classes of this component.

Figure 8. Package diagram of the Sentiment identifier module
Figura 8. Diagrama de paquetes del módulo Identificador de sentimientos



Source: Authors. Other packages

An object of the *AdminSA* class has access to the opinions that are stored in a list of *OutPLN* objects. Each *OutPLN* object represents a disaggregated opinion with each word, lemma and label. Additionally, the *AdminSa* object has access to the list of aspects of the previous process stored in a list of *Outcome* objects.

To identify opinion expressions, an object of the *SearchSentiment* class is created in the *PolarityClass* class. The

SearchSentiment class is in charge of finding the expressions of opinion related to each explicit aspect identified in the previous module. For this, the created object of the *SearchSentiment* class invokes the *obtainExprOpinion* method which has as input the list of *Outcome* objects (aspects) and a list of *OutPLN* objects (opinions).

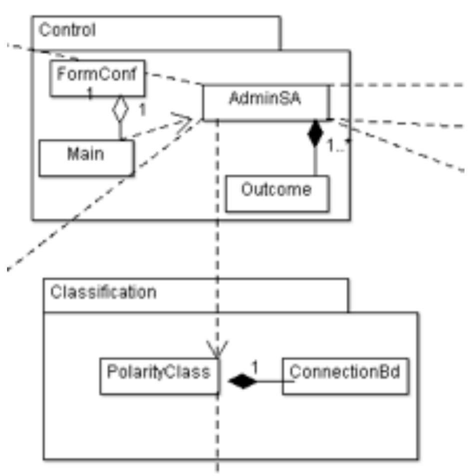
What is first done is to look for the position that has the aspect in the opinion that contains it. Then with each opinion (each *OutPLN* object) the nominal expressions are searched with adverb and adjective tags (defined experimentally) that are on a sliding window of variable length. For this system, a window length equal to two (2) was selected according to the experimental results made in the restaurant domain.

Additionally, in the search process of the expressions of each aspect, it is detected if the opinion is affected by denial or attenuation through grammatical rules. At this point you have a list of explicit aspects together with your set of expressions of opinion and a list of implicit aspects related to its explicit aspect. All these aspects are stored in the object collection of the Outcome class.

4.4 Sentiment classifier module

In this module, the polarity (positive, negative or neutral) of each of the explicit and implicit aspects is determined taking into account the expressions of opinion found in the previous module. Figure 9 shows the package diagram that includes the relationships between the classes of this component.

Figure 9. Package diagram of the Sentiment classifier module
Figura 9. Diagrama de paquetes del módulo Clasificador de sentimientos



Source: Authors. Other packages

The *AdminSA* class creates an object of the *PolarityClass* class to invoke the *calPolarity()* method which has as input the complete set of aspects in *OutCome* and an object of the *ConnectionBD* class which gives access to the domain corpus.

Initially, the explicit aspects are identified from each *OutCome* object. From each explicit aspect the expressions of opinion are obtained and together with the words sentiment seeds ***(****"excelente", "bueno", "malo", "pésimo", "indiferente")*, the *pmiAspExpOp* method is invoked which will yield a result of PMI for each aspect, expression of opinion and seed. If the returned PMI values are irregular (PMI <= 0) then the method *pmiAspPocoFrec* is invoked only with the expressions of opinion and seeds. Finally, the highest PMI value of each seed is imposed and it is the one that determines the polarity. If the highest PMI corresponds to the "*excelente*" and "*bueno*" seeds, the polarity is positive. If the highest PMI is of the "*malo*" and "*pesimo*" seeds, the polarity is negative. Otherwise, the polarity would be neutral.

The results of the polarity of each aspect are stored in Outcome, leaving the output of the system complete as a list of objects containing aspects, expressions of opinion and polarity. Finally, the negation and management methods are invoked to change the aspect polarity if the NEGATION and ATENUATION variables have been activated.

For the evaluation of the system, the results of experimentation carried out that focus on valuing the extraction of aspects and the sentiment classification are recorded. The measures to be used are precision, recall, F1 and accuracy. A series of experiments was carried out taking as reference the corpus of the task 5 related to Aspect-Based Sentiment Analysis of the 2016 edition of *SemeEval* (International Workshop on Semantic Evaluation) an organization that performs, as a competence , continuous evaluations of computational systems of semantic analysis. Specifically, sub-task 1 (SB1) was addressed in the restaurant domain in Spanish [24]. The following subtasks have been addressed from *Semeval*: the subtask corresponding to slot2 (extraction of aspects) and the subtask corresponding to slot3 (sentiment classification). For this, the corpus of the task consisting of 2070 training sentences and 881 evaluation sentences has been used. As an evaluation measure for slot2 the measure F1 was used and for slot3, the polarity, the measure that was used was accuracy.

The results of the task of extracting aspects and sentiment classification of the *AspectSA* system on the domain of restaurants in the evaluation corpus for *Semeval* tasks are shown in Table 1.

Table1. Results for extraction of aspects and sentiment classification
Tabla 1. Resultados para extracción de aspectos y clasificación de sentimientos

Process	Variable	Measured Value
Extraction of aspects	F1	73.07
Sentiment classification	Accuracy	84.8

The results obtained were compared with the final results of *Semeval* 2016 for the restaurant domain, subtask SB1 and Spanish language (see Table 2).

Table2. Results for extraction of aspects and sentiment classification
Tabla 2. Resultados de subtarea SB1 de Semeval

Lang. / Dom. / Sub.	Slot2 / F1	Slot3 / Acc.
SP	GTI/C/68.515	IIT-T./U/83.582
REST	GTI/U/68.387	TGB/C/82.09
SB1	IIT-T./U/64.338	UWB/C/81.343
	TGB/C/55.764	INSIG./C/79.571
	basel./C/51.914	basel./C/77.799

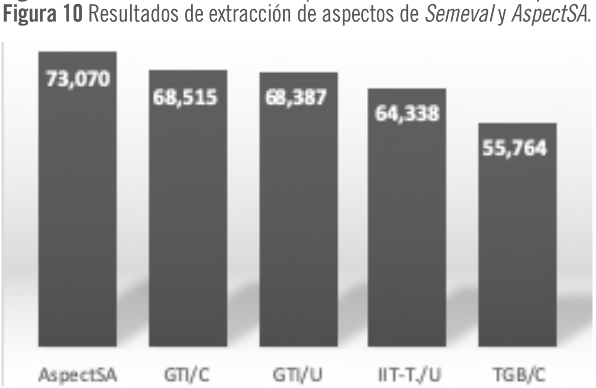
Table 2 shows a list by column of all participants in the competition only in sub-task 1 (SB1), in the restaurant domain (REST) and in the Spanish language (SP). In the list appears the name of the team followed by the letter U or C and then the value of the measurement. The letter C indicates that it is restricted only to the training data provided and the letter U indicates unrestricted which allows the use of additional resources, such as lexical or training data. The table shows the values of measure F1 for the first task and the measure of accuracy for the last task. In the final part of each list, the baseline is shown as the initial reference value. Table 3 shows the comparison between the results of the proposed system (*AspectSA*) and the results of the winners of the *Semeval* competition.

Table3. Results of comparison between *Semeval* and the proposed system.
Tabla 3. Resultados de comparación entre *Semeval* y el sistema propuesto.

Process	Variable	AspectSA	Semeval
Extraction of aspects	F1	73.07	68.51
Sentiment classification	Accuracy	84.8	83.58

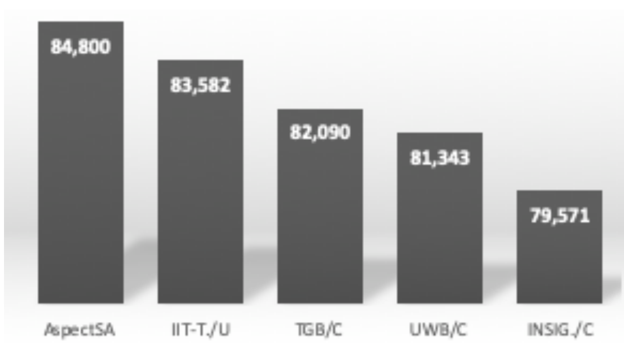
As can be seen, AspectSA obtains F1 values higher than the winners of the competition in the extraction of aspects. Analyzing the results, it should be noted that the choice and use of domain ontology was vital for the identification of aspects, since they represent the concepts of a given domain and its relationships, that is, they are an abstract model of a domain, where the concepts used are clearly defined and are not simple dictionaries. By reusing a domain ontology validated in other tasks, it was possible to perform an extraction that took into account the meaning because it was arranged in a specific domain taking advantage of the classes, individuals and relationships, this allowed to exploit this knowledge of the domain to improve the performance in the extraction of aspects. In Figure 10 you can see the results of extraction of aspects of the Semeval competition and the AspectSA system. Here only four (4) teams participated with scores between 55.76 and 68.51 of F1. Here the AspectSA system greatly exceeds the best system of the competition by almost 5 points.

Figure 10. Results of extraction of aspects of the *Semeval* and *AspectSA*
Figura 10 Resultados de extracción de aspectos de *Semeval* y *AspectSA.*



Analyzing polarity results, the proposed system achieves better results than those presented in *Semeval* obtained by the *IIT-T* team [34]. It should be noted that *AspectSA* works on an unsupervised approach that does not depend on the domain and does not work with tagged data compared to Kumar who needs a tagged sentiment lexicon for the task. It is important to note that the system only needs the domain corpus, the expressions of opinion and the seeds to obtain the sentiment associated. The larger the corpus, the system may find more relationships between aspects with the words of opinion, but this causes a drop in performance, so it was decided to work with the corpus of 40,000 opinions. Each time an opinion is processed, it is saved in the corpus, which allows the adjusted values of PMIPA to be adjusted little by little. In Figure 11 you can see the results of slot3 of the *Semeval* competition and the *AspectSA* system. Here only four (4) teams participated with scores between 79.57 and 83.58 of accuracy. Here the *AspectSA* system exceeds the best system of the competition by almost 1 point.

Figure 11. Results of Sentiment classification of the *Semeval* and *AspectSA*
Figura 11. Resultados de la clasificación de Sentimiento de *Semeval* y *AspectSA.*



It is important to highlight the advantages of the *AspectSA* system with respect to the other systems that currently work for the Spanish language. Initially, *AspectSA* is one of the few systems currently in existence that completely carries out the process of sentiment analyzing at the level of aspects totally focused on Spanish. It is also a completely unsupervised system that minimizes the human presence for the two main processes, the extraction of aspects and the sentiment classification. This allows the system to be more quickly scalable to any language or domain.

CONLUSION

In this paper a system for sentiment analysis at the level of aspects in Spanish (*AspectSA*) based on semantic techniques and measures of association was presented.

AspectSA is one of the few systems that completely carries out the process of sentiment analyzing at the level of aspects totally focused on Spanish. That is, it extracts explicit and implicit aspects, takes into account attenuation and negation, and determines the polarity of the aspects.

The system is completely unsupervised which allows it to be scalable to any language and domain. This was built using external libraries such as *Freeling* and MCR *WordNet*.

AspectSA was validated from the metrics used in the literature for these systems, as a measure -F1 and Precision, obtaining values of 73.07 of F1 in extraction of aspects and 84.8 of accuracy in sentiment classification, surpassing similar systems.

REFERENCES

[1] C. Henríquez and J. Guzmán, "Extracción de información desde la web para identificar acciones de un modelo de dominio en planificación automática.," *Ingeniare*, vol. 23, no. 3, pp. 439–448, 2015.

[2] C. Henríquez and J. A. Guzmán, "Modelo de extracción de información desde recursos web para aplicaciones de la planificación automática," *Prospectiva*, vol. 10, no. 2, pp. 74–80, 2012.

[3] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions*. 2015.

[4] C. Henríquez, J. Guzmán, and D. Salcedo, "Minería de Opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles Opinion," *Proces. del Leng. Nat.*, vol. 41, pp. 25–32, 2016.

[5] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 12, pp. 1–135, 2008.

[6] C. Henríquez, F. Plà, L. F. Hurtado, and J. A. G. Luna, "Análisis de sentimientos a nivel de aspecto usando ontologías y aprendizaje automático," *Proces. del Leng. Nat.*, vol. 59, pp. 49–56, 2017.

[7] J. Steinberger, T. Brychcín, and M. Konkol, "Aspect-Level Sentiment Analysis in Czech," in *Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2014, pp. 24–30.

[8] F. Ali, D. Kwak, P. Khan, S. M. R. Islam, K. H. Kim, and K. S. Kwak, "Fuzzy Ontology-based Sentiment Analysis of Transportation and City Feature Reviews for Safe Traveling," *Transp. Res. Part C Emerg. Technol.*, vol. 77, pp. 33–48, 2017.

[9] R. Hoogervorst *et al.*, "Aspect-Based Sentiment Analysis on the Web using Rhetorical Structure Theory," *Int. Conf. Web Eng. Springer Int.*, 2016.

[10] T. A. Rana and Y.-N. Cheah, "Aspect extraction in sentiment analysis: comparative analysis and survey," *Artif. Intell. Rev.*, vol. 46, no. 4, pp. 459–483, 2016.

[11] C. Henríquez and J. Guzmán, "A Review of Sentiment Analysis in Spanish," *Tecciencia*, vol. 12, no. 22, pp. 35–48, 2017.

[12] L. Sun, S. Li, J. Li, and J. Lv, "A novel context-based implicit feature extracting method," in *Data Science and Advanced Analytics (DSAA), 2014 International Conference on*, 2014, pp. 420–424.

[13] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, 2014.

[14] K. Schouten and F. Frasincar, "Finding Implicit Features in Consumer Reviews for Sentiment Analysis," in *Web Engineering: 14th International Conference, ICWE 2014, Toulouse, France, July 1-4, 2014. Proceedings*, S. Casteleyn, G. Rossi, and M. Winckler, Eds. Cham: Springer International Publishing, 2014, pp. 130–144.

[15] I. Peñalver-Martinez *et al.*, "Feature-based opinion mining through ontologies," *Expert Syst. with Appl.* , vol. 41, no. 13, pp. 5995–6008, 2014.

[16] H. Wang, Y. Lu, and C. Zhai, "Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach," 2010.

[17] W. Zhang, H. Xu, and W. Wan, "Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis," *Expert Syst. Appl.*, 2012.

[18] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion Word Expansion and Target Extraction through Double Propagation," *Comput. Linguist.*, vol. 37, pp. 9–27, 2011.

[19] D. Marcheggiani, O. Täckström, A. Esuli, and F. Sebastiani, "Hierarchical multi-label conditional random fields for aspect-oriented opinion mining," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014.

[20] F. Xianghua, L. Guo, G. Yanyan, and W. Zhiqiang, "Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon," *Knowledge-Based Systems*. 2013.

[21] E. Cambria, "An Introduction to Concept-Level Sentiment Analysis.," in *MICAI (2)*, 2013, pp. 478–483.

[22] K. Schouten and F. Frasincar, "Survey on Aspect-Level Sentiment Analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, 2016.

[23] L.-F. Hurtado and F. Pla, "ELiRF-UPV at TASS 2016: Sentiment Analysis in Twitter," in *TASS 2016: Workshop on Sentiment Analysis at SEPLN*, 2016, pp. 47–51.

[24] M. Pontiki *et al.*, "SemEval-2016 Task 5: Aspect Based Sentiment Analysis," in *Semeval*, 2016, pp. 19–30.

[25] L. A. De Freitas and R. Vieira, "Ontology-based Feature Level Opinion Mining for Portuguese Reviews," in *Proceedings of the 22nd International Conference on World Wide Web. ACM,* 2013, pp. 367–370.

[26] A. S. Manek, P. D. Shenoy, and M. C. Mohan, "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier," *World Wide Web*, pp. 1–20, 2016.

[27] S. and V. P. and K. G. and C. E. and Z. R. Hazarika Devamanyuand Poria, "Modeling Inter-Aspect Dependencies for Aspect-Based Sentiment Analysis," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)* , 2018, pp. 266–270.

[28] L. A. Jiménez-Zafra, S. M., Martín-Valdivia, M. T., Martínez-Cámara, E., & Ureña-López, "Combining resources to improve unsupervised sentiment analysis at aspect-level," *J. Inf. Sci.*, 2015.

[29] C. Wu, F. Wu, S. Wu, Z. Yuan, and Y. Huang, "A hybrid unsupervised method for aspect term and opinion target extraction," *Knowledge-Based Syst.*, vol. 148, pp. 66–73, May 2018.

[30] A. García-Pablos, M. Cuadros, and G. Rigau, "W2vlda: almost unsupervised system for aspect based sentiment analysis," *Expert Syst. Appl.*, vol. 91, pp. 127–137, 2018.

[31] L. Padró and E. Stanilovsky, "FreeLing 3.0: Towards Wider Multilinguality," in *LREC2012*, 2012.

[32] A. Gonzalez-Agirre and G. Rigau, "Construcción de una base de conocimiento léxico multilíngüe de amplia cobertura: Multilingual Central Repository Building a wide coverage multilingual lexical knowledge base: Multilingual Central Repository," *Linguamatica*, vol. 5, no. 1, pp. 13–28, 2013.

[33] K. W. Church and P. Hanks, "Word association norms, mutual information, and lexicography," *Comput. Linguist.*, vol. 16, no. 1, pp. 22–29, 1990.

[34] A. Kumar, S. Kohail, A. Kumar, A. Ekbal, and C. Biemann, "IIT-TUDA at SemEval-2016 Task 5: Beyond Sentiment Lexicon: Combining Domain Dependency and Distributional Semantics Features for Aspect Based Sentiment Analysis," in *Proceedings of SemEval*, 2016, pp. 1129–1135.