# UMCC_DLSI-(SA): Using a ranking algorithm and informal features to solve Sentiment Analysis in Twitter

**Yoan Gutiérrez, Andy González, Roger Pérez, José I. Abreu**
University of Matanzas, Cuba
{yoan.gutierrez, roger.perez ,jose.abreu}@umcc.cu, andy.gonzalez@infonet.umcc.cu

**Antonio Fernández Orquín, Alejandro Mosquera, Andrés Montoyo, Rafael Muñoz**
University of Alicante, Spain
antonybr@yahoo.com, {amosquera, montoyo, rafael}@dlsi.ua.es

**Franc Camara**
Independent Consultant
USA
info@franccamara .com

## Abstract

In this paper, we describe the development and performance of the supervised system UMCC_DLSI-(SA). This system uses corpora where phrases are annotated as Positive, Negative, Objective, and Neutral, to achieve new sentiment resources involving word dictionaries with their associated polarity. As a result, new sentiment inventories are obtained and applied in conjunction with detected informal patterns, to tackle the challenges posted in Task 2b of the Semeval-2013 competition. Assessing the effectiveness of our application in sentiment classification, we obtained a 69% F-Measure for neutral and an average of 43% F-Measure for positive and negative using Tweets and SMS messages.

## 1 Introduction

Textual information has become one of the most important sources of data to extract useful and heterogeneous knowledge from. Texts can provide factual information, such as: descriptions, lists of characteristics, or even instructions to opinion-based information, which would include reviews, emotions, or feelings. These facts have motivated dealing with the identification and extraction of opinions and sentiments in texts that require special attention.

Many researchers, such as (Balahur *et al.*, 2010; Hatzivassiloglou *et al.*, 2000; Kim and Hovy, 2006; Wiebe *et al.*, 2005) and many others have been working on this and related areas.

Related to assessment Sentiment Analysis (SA) systems, some international competitions have taken place. Some of those include: Semeval-2010 (Task 18: Disambiguating Sentiment Ambiguous Adjectives [1]) NTCIR (*Multilingual Opinion Analysis Task* (MOAT[2])) TASS[3] (Workshop on Sentiment Analysis at SEPLN workshop) and Semeval-2013 (Task 2[4] Sentiment Analysis in Twitter) (Kozareva *et al.*, 2013).

In this paper, we introduce a system for Task 2 b) of the Semeval-2013 competition.

### 1.1 Task 2 Description

In participating in "Task 2: Sentiment Analysis in Twitter" of Semeval-2013, the goal was to take a given message and its topic and classify whether it had a positive, negative, or neutral sentiment towards the topic. For messages conveying, both a positive and negative sentiment toward the topic, the stronger sentiment of the two would end up as the classification. Task 2 included two sub-tasks. Our team focused on Task 2 b), which provides two training corpora as described in Table 3, and two test corpora: 1) sms-test-input-B.tsv (with 2094 SMS) and 2) twitter-test-input-B.tsv (with 3813 Twit messages).

The following section shows some background approaches. Subsequently, in section 3, we describe the UMCC_DLSI-(SA) system that was used in Task 2 b). Section 4 describes the assessment of the obtained resource from the Sentiment Classification task. Finally, the conclusion and future works are presented in section 5.

## 2 Background

The use of sentiment resources has proven to be a necessary step for training and evaluating systems that implement sentiment analysis, which also

[1] http://semeval2.fbk.eu/semeval2.php
[2] http://research.nii.ac.jp/ntcir/ntcir-ws8/meeting/
[3] http://www.daedalus.es/TASS/
[4] http://www.cs.york.ac.uk/semeval-2013/task2/



*Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 443–449, Atlanta, Georgia, June 14-15, 2013. 

include fine-grained opinion mining (Balahur, 2011).

In order to build sentiment resources, several studies have been conducted. One of the first is the relevant work by (Hu and Liu, 2004) using lexicon expansion techniques by adding synonymy and antonym relations provided by WordNet (Fellbaum, 1998; Miller *et al.*, 1990) Another one is the research described by (Hu and Liu, 2004; Liu *et al.*, 2005) which obtained an Opinion Lexicon compounded by a list of positive and negative opinion words or sentiment words for English (around 6800 words).

A similar approach has been used for building WordNet-Affect (Strapparava and Valitutti, 2004) which expands six basic categories of emotion; thus, increasing the lexicon paths in WordNet.

Nowadays, many sentiment and opinion messages are provided by Social Media. To deal with the informalities presented in these sources, it is necessary to have intermediary systems that improve the level of understanding of the messages. The following section offers a description of this phenomenon and a tool to track it.

### 2.1 Text normalization

Several informal features are present in opinions extracted from Social Media texts. Some research has been conducted in the field of lexical normalization for this kind of text. TENOR (Mosquera and Moreda, 2012) is a multilingual text normalization tool for Web 2.0 texts with an aim to transform noisy and informal words into their canonical form. That way, they can be easily processed by NLP tools and applications. TENOR works by identifying out-of-vocabulary (OOV) words such as slang, informal lexical variants, expressive lengthening, or contractions using a dictionary lookup and replacing them by matching formal candidates in a word lattice using phonetic and lexical edit distances.

### 2.2 Construction of our own Sentiment Resource

Having analyzed the examples of SA described in section 2, we proposed building our own sentiment resource (Gutiérrez *et al.*, 2013) by adding lexical and informal patterns to obtain classifiers that can deal with Task 2b of Semeval-2013. We proposed the use of a method named RA-SR (using Ranking Algorithms to build Sentiment Resources) (Gutiérrez *et al.*, 2013) to build sentiment word inventories based on senti-semantic evidence obtained after exploring text with annotated sentiment polarity information. Through this process, a graph-based algorithm is used to obtain auto-balanced values that characterize sentiment polarities, a well-known technique in Sentiment Analysis. This method consists of three key stages: **(I)** Building contextual word graphs; **(II)** Applying a ranking algorithm; and **(III)** Adjusting the sentiment polarity values.

These stages are shown in the diagram in Figure 1, which the development of sentimental resources starts off by giving four corpora of annotated sentences (the first with neutral sentences, the second with objective sentences, the third with positive sentences, and the last with negative sentences).

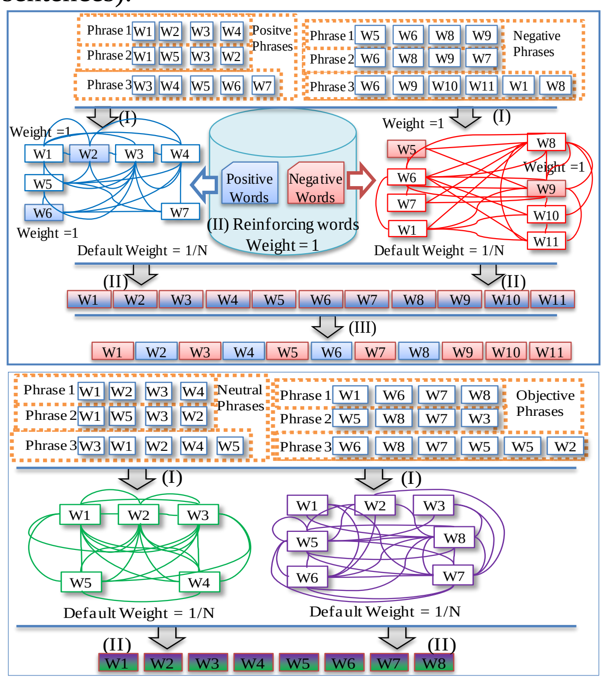


Figure 1. Resource walkthrough development process.

### 2.3 Building contextual word graphs

Initially, text preprocessing is performed by applying a Post-Tagging tool (using Freeling (Atserias *et al.*, 2006) tool version 2.2 in this case) to convert all words to lemmas[5]. After that, all obtained lists of lemmas are sent to RA-SR, then divided into four groups: neutral, objective, positive, and negative candidates. As the first set

[5] Lemma denotes canonic form of the words.

of results, four contextual graphs are obtained: $Gneu,\ Gobj$, $Gpos$, and $Gneg$, where each graph includes the words/lemmas from the neutral, objective, positive and negative sentences respectively. These graphs are generated after connecting all words for each sentence into individual sets of annotated sentences in concordance with their annotations ($Pos$, $Neg$, $Obj$, $Neu$ ).

Once the four graphs representing neutral, objective, positive and negative contexts are created, we proceed to assign weights to apply graph-based ranking techniques in order to auto-balance the particular importance of each vertex $v_i$ into $Gneu$, $Gobj$, $Gpos$ and $Gneg$.

As the primary output of the graph-based ranking process, the positive, negative, neutral, and objective values are calculated using the PageRank algorithm and normalized with equation (1). For a better understanding of how the contextual graph was built see (Gutiérrez *et al.*, 2013).

## 2.4 Applying a ranking algorithm

To apply a graph-based ranking process, it is necessary to assign weights to the vertices of the graph. Words involved into $Gneu,\ Gobj, Gpos$ and $Gneg$ take the default of $1/\mathrm{N}$ as their weight to define the weight of $v$ vector, which is used in our proposed ranking algorithm. In the case where words are identified on the sentiment repositories (see Table 4) as positive or negative, in relation to their respective graph, a weight value of 1 (in a range $[0 \ldots 1]$ ) is assigned. $N$ represents the maximum quantity of words in the current graph. After that, a graph-based ranking algorithm is applied in order to structurally raise the graph vertexes' voting power. Once the reinforcement values are applied, the proposed ranking algorithm is able to increase the significance of the words related to these empowered vertices.

The PageRank (Brin and Page, 1998) adaptation, which was popularized by (Agirre and Soroa, 2009) in Word Sense Disambiguation thematic, and which has obtained relevant results, was an inspiration to us in our work. The main idea behind this algorithm is that, for each edge between $v_i$ and $v_j$ in graph $G$, a vote is made from $v_i$ to $v_j$. As a result, the relevance of $v_j$ is increased.

On top of that, the vote strength from $i$ to $j$ depends on $v_i's$ relevance. The philosophy behind it is that, the more important the vertex is, the more strength the voter would have. Thus, PageRank is generated by applying a random walkthrough from the internal interconnection of $G$, where the final relevance of $v_i$ represents the random walkthrough probability over $G$, and ending on $v_i$.

In our system, we apply the following configuration: dumping factor $c = 0.85$ and, like in (Agirre and Soroa, 2009) we used 30 iterations. A detailed explanation about the PageRank algorithm can be found in (Agirre and Soroa, 2009)

After applying PageRank, in order to obtain standardized values for both graphs, we normalize the rank values by applying the equation (1), where $Max(\mathbf{Pr})$ obtains the maximum rank value of $\boldsymbol{Pr}$ vector (rankings' vector).

$$\mathbf{Pr}_i = \mathbf{Pr}_i / Max(\mathbf{Pr}) \qquad (1)$$

## 2.5 Adjusting the sentiment polarity values

After applying the PageRank algorithm on$Gneu$, $Gobj$, $Gpos$ and $Gneg$, having normalized their ranks, we proceed to obtain a final list of lemmas (named $Lf$ ) while avoiding repeated elements. $Lf$ is represented by $Lf_i$ lemmas, which would have, at that time, four assigned values: Neutral, Objective, Positive, and Negative, all of which correspond to a calculated rank obtained by the PageRank algorithm.

At that point, for each lemma from $Lf$, the following equations are applied in order to select the definitive subjectivity polarity for each one:

$$Pos = \begin{cases} Pos - Neg\ ; & Pos > Neg \\ 0 & ; otherwise \end{cases} \qquad (2)$$

$$Neg = \begin{cases} Neg - Pos\ ; & Neg > Pos \\ 0 & ; otherwise \end{cases} \qquad (3)$$

Where $Pos$ is the Positive value and $Neg$ the Negative value related to each lemma in $Lf$.

In order to standardize again the $Pos$ and $Neg$ values and making them more representative in a $[0 \ldots 1]$ scale, we proceed to apply a normalization process over the $Pos$ and $Neg$ values.

From there, based on the objective features commented by (Baccianella *et al.*, 2010), we assume the same premise to establish an alternative objective value of the lemmas. Equation (4) is used for that:

$$ObjAlt = 1 - |Pos - Neg| \qquad (4)$$

Where $ObjAlt$ represents the alternative objective value.

As a result, each word obtained in the sentiment resource has an associated value of: positivity ($Pos$, see equation (2)), negativity ($Neg$, see equation (3)), objectivity ($Real\_obj$, obtained by PageRank over $Gobj$ and normalized with equation (1)), calculated-objectivity ($ObjAlt$, now cited as $obj\_measured$) and neutrality ($Neu$, obtained by PageRank over $Gneu$ and normalized with equation (1)).

## 3 System Description

The system takes annotated corpora as input from which two models are created. One model is created by using only the data provided at Semeval-2013 (Restricted Corpora, see Table 3), and the other by using extra data from other annotated corpora (Unrestricted Corpora, see Table 3). In all cases, the phrases are pre-processed using Freeling 2.2 pos-tagger (Atserias *et al.*, 2006) while a dataset copy is normalized using TENOR (described in section 2.1).

The system starts by extracting two sets of features. The Core Features (see section 3.1) are the Sentiment Measures and are calculated for a standard and normalized phrase. The Support Features (see section 3.2) are based on regularities, observed in the training dataset, such as emoticons, uppercase words, and so on.

The supervised models are created using Weka[6] and a Logistic classifier, both of which the system uses to predict the values of the test dataset. The selection of the classifier was made after analyzing several classifiers such as: Support Vector Machine, J48 and REPTree. Finally, the Logistic classifier proved to be the best by increasing the results around three perceptual points.

The test data is preprocessed in the same way the previous corpora were. The same process of feature extraction is also applied. With the aforementioned features and the generated models, the system proceeds to classify the final values of Positivity, Negativity, and Neutrality.

### 3.1 The Core Features

The Core Features is a group of measures based on the resource created early (see section 2.2). The system takes a sentence preprocessed by Freeling 2.2 and TENOR. For each lemma of the analyzed sentence, $pos$, $neg$, $obj\_measured$, $real\_obj$, and $neu$ are calculated by using the respective word values assigned in RA-SR. The obtained values correspond to the sum of the corresponding values for each intersecting word between the analyzed sentence (lemmas list) and the obtained resource by RA-SR. Lastly, the aforementioned attributes are normalized by dividing them by the number of words involved in this process.

Other calculated attributes are: $pos\_count$, $neg\_count$, $obj\_measured\_count$, $obj\_real\_count$ and $neu\_count$. These attributes count each involved iteration for each feature type ($Pos$, $Neg$, $Real\_obj$, $ObjAlt$ and $Neu$ respectively, where the respective value may be greater than zero.

Attributes *cnp* and *cnn* are calculated by counting the amount of lemmas in the phrases contained in the Sentiment Lexicons (Positive and Negative respectively).

All of the 12 attributes described previously are computed for both, the original, and the normalized (using TENOR) phrase, totaling 24 attributes. The Core features are described next.

| Feature Name | Description |
|---|---|
| ***pos*** | Sum of respective value of each word. |
| ***neg*** | |
| ***obj_measured*** | |
| ***real_obj*** | |
| ***neu*** | |
| ***pos_count*** | Counts the words where its respective value is greater than zero |
| ***neg_count*** | |
| ***obj_measured_count*** | |
| ***real_obj_count*** | |
| ***neu_count*** | |
| ***cnp*** (to positive) | Counts the words contained in the Sentiment Lexicons for their respective polarities. |
| ***cnn*** (to negative) | |

Table 1. Core Features

### 3.2 The Support Features

The Support Features is a group of measures based on characteristics of the phrases, which may help with the definition on extreme cases. The *emotPos* and *emotNeg* values are the amount of Positive and Negative Emoticons found in the phrase. The *exc* and *itr* are the amount of exclamation and interrogation signs in the phrase. The following table shows the attributes that represent the support features:

| Feature Name | Description |
|---|---|
| ***emotPos*** | Counts the respective Emoticons |
| ***emotNeg*** | |
| ***exc*** (exclamation marks ("!")) | Counts the respective marks |
| ***itr*** (question marks ("?")) | |
| ***WORDS_count*** | Counts the uppercase words |
| ***WORDS_pos*** | Sums the respective values of the Uppercase words |
| ***WORDS_neg*** | |
| ***WORDS_pos_count_res*** (to | Counts the Uppercase words |

[6] http://www.cs.waikato.ac.nz/

| | |
|---|---|
| positivity) | contained in their respective Graph |
| $\boldsymbol{WORDS\_neg\_count\_res}$(to negativity) | |
| $\boldsymbol{WORDS\_pos\_count\_dict}$ (to positivity) | Counts the Uppercase words contained in the Sentiment Lexicons [7] for their respective polarity |
| $\boldsymbol{WORDS\_neg\_count\_dict}$ (to negativity) | |
| $\boldsymbol{wooords\_count}$ | Counts the words with repeated chars |
| $\boldsymbol{wooords\_pos}$ | Sums the respective values of the words with repeated chars |
| $\boldsymbol{wooords\_neg}$ | |
| $\boldsymbol{wooords\_neg\_count\_dict}$ (in negative lexical resource ) | Counts the words with repeated chars contained in the respective lexical resource |
| $\boldsymbol{wooords\_pos\_count\_dict}$ (in positive lexical resource ) | |
| $\boldsymbol{wooords\_pos\_count\_res}$ (in positive graph ) | Counts the words with repeated chars contained in the respective graph |
| $\boldsymbol{wooords\_neg\_count\_res}$ (in negative graph ) | |

Table 2. The Support Features

## 4 Evaluation

In the construction of the sentiment resource, we used the annotated sentences provided by the corpora described in Table 3. The resources listed in Table 3 were selected to test the functionality of the words annotation proposal with subjectivity and objectivity. Note that the shadowed rows correspond to constrained runs corpora: tweeti-b-sub.dist_out.tsv [8] (dist), b1_tweeti-objorneu-b.dist_out.tsv [9] (objorneu), twitter-dev-input-B.tsv[10] (dev).

The resources from Table 3 that include unconstrained runs corpora are: all the previously mentioned ones, Computational-intelligence[11] (CI) and stno[12] corpora.

The used sentiment lexicons are from the WordNetAffect_Categories[13] and opinion-words[14] files as shown in detail in Table 4.

Some issues were taken into account throughout this process. For instance, after obtaining a contextual graph $G$, factotum words are present in most of the involved sentences (i.e., verb "*to be").* This issue becomes very dangerous after applying the PageRank algorithm because the algorithm strengthens the nodes possessing many linked elements. For that reason, the subtractions $Pos - Neg$ and $Neg - Pos$ are applied, where the most frequent words in all contexts obtain high values. The subtraction becomes a dumping factor.

As an example, when we take the verb "*to be*", before applying equation (1), the verb achieves the highest values in each subjective context graph ($Gpos$ and $Gneg$) namely, 9.94 and 18.67 rank values respectively. These values, once equation (1) is applied, are normalized obtaining both $Pos = 1$ and $Neg = 1$ in a range [0...1]. At the end, when the following steps are executed (Equations (2) and (3)), the verb "*to be*" achieves $Pos = 0$ , $Neg = 0$ and therefore $ObjAlt = 1$. Through this example, it seems as though we subjectively discarded words that appear frequently in both contexts (Positive and Negative).

| **Corpus** | **N** | **P** | **O** | **Neu** | **Obj or Neu** | **Unk** | **T** | **C** | **UC** |
|---|---|---|---|---|---|---|---|---|---|
| dist | 176 | 368 | 110 | 34 | - | - | 688 | X | X |
| objorneu | 828 | 1972 | 788 | 1114 | 1045 | - | 5747 | X | X |
| dev | 340 | 575 | - | 739 | - | - | 1654 | X | X |
| CI | 6982 | 6172 | - | - | - | - | 13154 | | X |
| stno[15] | 1286 | 660 | - | 384 | - | 10000 | 12330 | | X |
| T | 9272 | 9172 | 898 | 1532 | 1045 | 10000 | 31919 | | |

Table 3. Corpora used to apply RA-SR. Positive (P), Negative (N), Objective (Obj/O), Unknow (Unk), Total (T), Constrained (C), Unconstrained (UC).

| **Sources** | **P** | **N** | **T** |
|---|---|---|---|
| **WordNet-Affects_Categories** (Strapparava and Valitutti, 2004) | 629 | 907 | 1536 |
| **opinion-words** (Hu and Liu, 2004; Liu *et al.*, 2005) | 2006 | 4783 | 6789 |
| Total | 2635 | 5690 | 8325 |

Table 4. Sentiment Lexicons. Positive (P), Negative (N) and Total (T).

| | | | **Precision (%)** | | | **Recall (%)** | | | **Total (%)** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **C** | **Inc** | **P** | **N** | **Neu** | **P** | **N** | **Neu** | **Prec** | **Rec** | **F1** |
| Run1 | 8032 | 1631 | 80,7 | 83,8 | 89,9 | 90,9 | 69,5 | 86,4 | **84,8** | **82,3** | **82,9** |
| Run2 | 19101 | 4671 | 82,2 | 77,3 | 89,4 | 80,7 | 81,9 | 82,3 | **83,0** | **81,6** | **80,4** |

Table 5. Training dataset evaluation using cross-validation (Logistic classifier (using 10 folds)). Constrained (Run1), Unconstrained (Run2), Correct(C), Incorrect (Inc).

### 4.1 The training evaluation

In order to assess the effectiveness of our trained classifiers, we performed some evaluation tests. Table 5 shows relevant results obtained after applying our system to an environment (specific domain). The best results were obtained with the

[7] Resources described in Table 4.
[8] Semeval-2013 (Task 2. Sentiment Analysis in Twitter, subtask b).
[9] Semeval-2013 (Task 2. Sentiment Analysis in Twitter, subtask b).
[10] http://www.cs.york.ac.uk/semeval-2013/task2/
[11] A sentimental corpus obtained applying techniques developed by GPLSI department. See (http://gplsi.dlsi.ua.es/gplsi11/allresourcespanel)
[12] NTCIR Multilingual Opinion Analysis Task (MOAT) http://research.nii.ac.jp/ntcir/ntcir-ws8/meeting/
[13] http://wndomains.fbk.eu/wnaffect.html
[14] http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html
[15] NTCIR Multilingual Opinion Analysis Task (MOAT) http://research.nii.ac.jp/ntcir/ntcir-ws8/meeting/

restricted corpus. The information used to increase the knowledge was not balanced or perhaps is of poor quality.

### 4.2 The test evaluation

The test dataset evaluation is shown in Table 6, where system results are compared with the best results in each case. We notice that the constrained run is better in almost every aspect. In the few cases where it was lower, there was a minimal difference. This suggests that the information used to increase our Sentiment Resource was unbalanced (high difference between quantity of tagged types of annotated phrases), or was of poor quality. By comparing these results with the ones obtained by our system on the test dataset, we notice that on the test dataset, the results fell in the middle of the effectiveness scores. After seeing these results (Table 5 and Table 6), we assumed that our system performance is better in a controlled environment (or specific domain). To make it more realistic, the system must be trained with a bigger and more balanced dataset.

Table 6 shows the results obtained by our system while comparing them to the best results of Task 2b of Semeval-2013. In Table 5, we can see the difference between the best systems. They are the ones in bold and underlined as target results. These results have a difference of around 20 percentage points. The grayed out ones correspond to our runs.

| | | | **Precision (%)** | | | **Recall (%)** | | | **Total** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Runs** | **C** | **Inc** | **P** | **N** | **Neu** | **P** | **N** | **Neu** | **Prec** | **Rec** | **F 1** |
| 1_tw | 2082 | 1731 | 60,9 | 46,5 | 52,8 | 49,8 | 41,4 | 64,1 | 53,4 | 51,8 | **49,3** |
| 1_tw_cnd | 2767 | 1046 | 81,4 | 69,7 | 67,7 | 66,7 | 60,4 | 82,6 | 72,9 | 69,9 | 69,0 |
| 2_tw | 2026 | 1787 | 58,0 | 42,2 | 42,2 | 52,2 | 43,9 | 57,4 | 47,4 | 51,2 | 49,0 |
| 2_tw_ter | 2565 | 1248 | 71,1 | 54,6 | 68,6 | 74,7 | 59,4 | 63,1 | 64,8 | 65,7 | 64,9 |
| 1_sms | 1232 | 862 | 43,9 | 46,1 | **69,5** | 55,9 | 31,7 | **68,9** | 53,2 | 52,2 | **43,4** |
| 1_sms_cnd | 1565 | 529 | 73,1 | 55,4 | 85,2 | 73,0 | 75,4 | 75,3 | 71,2 | 74,5 | 68,5 |
| 2_sms | 1023 | 1071 | 38,4 | 31,4 | 68,3 | 60,0 | 38,3 | 47,8 | 46,0 | 48,7 | 40,7 |
| 2_sms_ava | 1433 | 661 | 60,9 | 49,4 | 81,4 | 65,9 | 63,7 | 71,0 | 63,9 | 66,9 | 59,5 |

Table 6. Test dataset evaluation using official scores. Corrects(C), Incorrect (Inc).

Table 6 run descriptions are as follows:

- UMCC_DLSI_(SA)-B-twitter-constrained (1_tw),
- NRC-Canada-B-twitter-constrained (1_tw_cnd),
- UMCC_DLSI_(SA)-B-twitter-unconstrained (2_tw),
- teragram-B-twitter-unconstrained (2_tw_ter),
- UMCC_DLSI_(SA)-B-SMS-constrained (1_sms),
- NRC-Canada-B-SMS-constrained (1_sms_cnd), UMCC_DLSI_(SA)-B-SMS-unconstrained (2_sms),
- AVAYA-B-sms-unconstrained (2_sms_ava).

As we can see in the training and testing evaluation tables, our training stage offered more relevant scores than the best scores in Task2b (Semaval-2013). This means that we need to identify the missed features between both datasets (training and testing).

For that reason, we decided to check how many words our system (more concretely, our Sentiment Resource) missed. Table 7 shows that our system missed around 20% of the words present in the test dataset.

| | **hits** | **miss** | **miss (%)** |
|---|---|---|---|
| twitter | 23807 | 1591 | 6,26% |
| sms | 12416 | 2564 | 17,12% |
| twitter nonrepeat | 2426 | 863 | **26,24%** |
| sms norepeat | 1269 | 322 | **20,24%** |

Table 7. Quantity of words used by our system over the test dataset.

## 5 Conclusion and further work

Based on what we have presented, we can say that we could develop a system that would be able to solve the SA challenge with promising results. The presented system has demonstrated election performance on a specific domain (see Table 5) with results over 80%. Also, note that our system, through the SA process, automatically builds sentiment resources from annotated corpora.

For future research, we plan to evaluate RA-SR on different corpora. On top of that, we also plan to deal with the number of neutral instances and finding more words to evaluate the obtained sentiment resource.


## Acknowledgments

This research work has been partially funded by the Spanish Government through the project TEXT-MESS 2.0 (TIN2009-13391-C04), "Análisis de Tendencias Mediante Técnicas de Opinión Semántica" (TIN2012-38536-C03-03) and “Técnicas de Deconstrucción en la Tecnologías del Lenguaje Humano” (TIN2012-31224); and by the Valencian Government through the project PROMETEO (PROMETEO/2009/199).

## References


Agirre, E. and A. Soroa. Personalizing PageRank for Word Sense Disambiguation. Proceedings of the 12th conference of the European chapter of the Association for Computational Linguistics (EACL-2009), Athens, Greece, 2009.

Atserias, J.; B. Casas; E. Comelles; M. González; L. Padró and M. Padró. FreeLing 1.3: Syntactic and semantic services in an opensource NLP library. Proceedings of LREC'06, Genoa, Italy, 2006.

Baccianella, S.; A. Esuli and F. Sebastiani. SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. 7th Language Resources and Evaluation Conference, Valletta, MALTA., 2010. 2200-2204 p.

Balahur, A. Methods and Resources for Sentiment Analysis in Multilingual Documents of Different Text Types. Department of Software and Computing Systems. Alacant, Univeristy of Alacant, 2011. 299. p.

Balahur, A.; E. Boldrini; A. Montoyo and P. Martinez-Barco. The OpAL System at NTCIR 8 MOAT. Proceedings of NTCIR-8 Workshop Meeting, Tokyo, Japan., 2010. 241-245 p.

Brin, S. and L. Page The anatomy of a large-scale hypertextual Web search engine Computer Networks and ISDN Systems, 1998, 30(1-7): 107-117.

Fellbaum, C. WordNet. An Electronic Lexical Database. University of Cambridge, 1998. p. The MIT Press.

Gutiérrez, Y.; A. González; A. F. Orquín; A. Montoyo and R. Muñoz. RA-SR: Using a ranking algorithm to automatically building resources for subjectivity analysis over annotated corpora. 4th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA 2013), Atlanta, Georgia, 2013.

Hatzivassiloglou; Vasileios and J. Wiebe. Effects of Adjective Orientation and Gradability on Sentence Subjectivity. International Conference on Computational Linguistics (COLING-2000), 2000.

Hu, M. and B. Liu. Mining and Summarizing Customer Reviews. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), USA, 2004.

Kim, S.-M. and E. Hovy. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In Proceedings of workshop on sentiment and subjectivity in text at proceedings of the 21st international conference on computational linguistics/the 44th annual meeting of the association for computational linguistics (COLING/ACL 2006), Sydney, Australia, 2006. 1-8 p.

Kozareva, Z.; P. Nakov; A. Ritter; S. Rosenthal; V. Stoyonov and T. Wilson. Sentiment Analysis in Twitter. in: Proceedings of the 7th International Workshop on Semantic Evaluation. Association for Computation Linguistics, 2013.

Liu, B.; M. Hu and J. Cheng. Opinion Observer: Analyzing and Comparing Opinions on the Web. Proceedings of the 14th International World Wide Web conference (WWW-2005), Japan, 2005.

Miller, G. A.; R. Beckwith; C. Fellbaum; D. Gross and K. Miller. Five papers on WordNet. Princenton University, Cognositive Science Laboratory, 1990.

Mosquera, A. and P. Moreda. TENOR: A Lexical Normalisation Tool for Spanish Web 2.0 Texts. in: Text, Speech and Dialogue - 15th International Conference (TSD 2012). Springer, 2012.

Strapparava, C. and A. Valitutti. WordNet-Affect: an affective extension of WordNet. Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, 2004. 1083-1086 p.

Wiebe, J.; T. Wilson and C. Cardie. Annotating Expressions of Opinions and Emotions in Language. Kluwer Academic Publishers, Netherlands, 2005.