TASS 2013 - Workshop on Sentiment Analysis at SEPLN 2013: An overview

TASS 2013 - Taller de Análisis de Sentimientos en la SEPLN 2013: Introducción

Julio Villena-Román
Janine García-Morera
Daedalus, S.A.
{jvillena, jgarcia}@daedalus.es

Resumen: Este artículo describe el desarrollo de TASS 2013, la segunda edición del taller de evaluación experimental en el contexto de la SEPLN para fomentar la investigación en el campo del análisis de sentimiento en los medios sociales, específicamente centrado en el idioma español. El principal objetivo es promover el diseño de nuevas técnicas y algoritmos y la aplicación de los ya existentes para la implementación de complejos sistemas capaces de realizar un análisis de sentimientos basados en opiniones de textos cortos extraídos de medios sociales (concretamente Twitter). Este artículo describe las tareas propuestas en la edición de 2013, el contenido, formato y las estadísticas más importantes de los corpus generados, la lista de participantes y los diferentes enfoques planteados, así como los resultados generales obtenidos.
Palabras clave: TASS 2013, análisis de reputación, análisis de sentimientos, medios sociales

Abstract: This paper describes TASS 2013, the second edition of an experimental evaluation workshop within SEPLN to foster the research in the field of sentiment analysis in social media, specifically focused on Spanish language. The main objective is to promote the application of existing state-of-the-art algorithms and techniques and the design of new ones for the implementation of complex systems able to perform a sentiment analysis based on short text opinions extracted from social media messages (specifically Twitter) published by representative personalities. The paper presents the proposed tasks in the 2013 edition, the contents, format and main statistics of the generated corpus, the participant groups and their different approaches, and, finally, the overall results achieved.
Keywords: TASS 2013, reputation analysis, sentiment analysis, social media.

1 Introduction

TASS is an experimental evaluation workshop for sentiment analysis and online reputation analysis focused on Spanish language, organized as a satellite event of the annual SEPLN Conference. After a successful first edition in 2012[1] (Villena *et al.*, 2013), TASS 2013[2] was held on September 20th, 2013 at Universidad Complutense de Madrid, Madrid, Spain.

The long-term objective of TASS is to foster research in the field of reputation analysis. Reputation analysis is the process of tracking, investigating and reporting an entity's actions and other entities' opinions about those actions. In turn, reputation (according to Merriam-Webster dictionary) is "the overall quality or character of a given person or organization as seen or judged by people in general", or, in other words, the general recognition by other people of some characteristics or abilities for a given entity. In business, reputation comprises the actions of a company and its internal stakeholders along with the perception of consumers about the business, and affects

[1] http://www.daedalus.es/TASS2012
[2] http://www.daedalus.es/TASS2013

attitudes like satisfaction, commitment and trust, and drives behavior like loyalty and support.

Reputation analysis has come into wide use as a major factor of competitiveness in the increasingly complex marketplace of personal and business relationships among people and companies. The rise of social media such as blogs and social networks and the increasing amount of user-generated contents in the form of reviews, recommendations, ratings and any other form of opinion, has led to creation of an emerging trend towards online reputation analysis, i.e., the use of technologies to calculate the reputation value of a given entity based on the opinions that people show in social media about that entity. All of them are becoming promising topics in the field of marketing and customer relationship management, as the social media and its associated word-of-mouth effect is turning out to be the most important source of information for companies and their customers' sentiments towards their brands and products.

As a first approach, reputation analysis has two technological aspects: sentiment analysis and text classification (or categorization).

Sentiment analysis is the application of natural language processing and text analytics to identify and extract subjective information from texts, Sentiment analysis is a major technological challenge. The task is so hard that even humans often disagree on the sentiment of a given text. The fact that issues that one individual finds acceptable or relevant may not be the same to others, along with multilingual aspects, cultural factors and different contexts make it very hard to classify a text written in a natural language into a positive or negative sentiment. And the shorter the text is, for example, when analyzing Twitter messages or short comments in Facebook, the harder the task becomes.

On the other hand, automatic text classification is used to guess the topic of the text, among those of a predefined set of categories or classes, so as to be able to assign the reputation level of the company into different facets, axis or points of view of analysis. Text classification techniques, although studied for a longer time, still need more research effort to be able to build complex models with many categories with less workload and increase the precision and recall of the results. In addition, these models should work well with short texts and deal with specific text features that are present in social media messages (such as spelling mistakes, abbreviations, SMS language, etc.).

Within this context, the aim of TASS is to provide a forum for discussion and communication where the latest research work and developments in the field of sentiment analysis in social media, specifically focused on Spanish language, can be shown and discussed by scientific and business communities. The main objective is to promote the application of existing state-of-the-art algorithms and techniques and the design of new ones for the implementation of complex systems able to perform a sentiment analysis and text classification on short text opinions extracted from social media messages (specifically Twitter) published by a series of representative personalities.

The setup is based on a series of challenge tasks that are intended to provide a benchmark forum for comparing the latest approaches in these fields. In addition, with the creation and release of the fully tagged corpus, we aim to provide a benchmark dataset that enables researchers to compare their algorithms and systems.

The rest of the paper is organized as follows. Section 2 describes the corpus provided to participants and used for the challenge tasks. The third section describes the different tasks proposed this edition. Section 4 describes the participants and the overall results are presented in Section 5. The last section draws some conclusions and future directions.

2 Corpus

TASS 2013 experiments will be based on two different corpus.

2.1 General corpus

The general corpus contains over 68 000 Twitter messages, written in Spanish by about 150 well-known personalities and celebrities of the world of politics, economy, communication, mass media and culture, between November 2011 and March 2012. Although the context of extraction has a Spain-focused bias, the diverse nationality of the authors, including people from Spain, Mexico, Colombia, Puerto Rico, USA and many other countries, makes the corpus reach a global coverage in the Spanish-speaking world, which may allow to perform experiments for instance on the usage of different varieties of Spanish by different users based on their geographical information.

Each Twitter message includes its ID (tweetid), the creation date (date) and the user ID (user). Due to restrictions in the Twitter API Terms of Service[3], it is forbidden to redistribute a corpus that includes text contents or information about users. However, it is valid if those fields are removed and instead IDs (including Tweet IDs and user IDs) are provided. The actual message content can be easily obtained by making queries to the Twitter API using the tweetid.

The general corpus has been divided into two sets: *training* (about 10%) and *test* (90%). The training set will be released so that participants may train and validate their models for classification and sentiment analysis. The test corpus will be provided without any tagging and will be used to evaluate the results provided by the different systems.

Each message in both the training and test set is tagged with its *global polarity*, indicating whether the text expresses a positive, negative or neutral sentiment, or no sentiment at all. 5 levels have been defined: strong positive (P+), positive (P), neutral (NEU), negative (N), strong negative (N+) and one additional no sentiment tag (NONE).

In addition, there is also an indication of the level of agreement or disagreement of the expressed sentiment within the content, with two possible values: AGREEMENT and DISAGREEMENT. This is especially useful to make out whether a neutral sentiment comes from neutral keywords or else the text contains positive and negative sentiments at the same time.

Moreover, the *polarity at entity level*, i.e., the polarity values related to the entities that are mentioned in the text, is also included for those cases when applicable. These values are similarly divided into 5 levels and include the level of agreement as related to each entity.

On the other hand, a selection of a set of *topics* has been made based on the thematic areas covered by the corpus, such as *politics*, *soccer*, *literature* or *entertainment*. Each message in both the training and test set has been assigned to one or several of these topics (most messages are associated to just one topic, due to the short length of the text).

[3] https://dev.twitter.com/terms/api-terms

All tagging has been done semi automatically: a baseline machine learning model is first run and then all tags are manually checked by human experts. In the case of the polarity at entity level, due to the high volume of data to check, this tagging has just been done for the training set.

Table 1 shows a summary of the training and test corpora provided to participants.

Attribute	Value
Tweets	68 017
Tweets (test)	60 798 (89%)
Tweets (test)	60 798 (11%)
Topics	10
Tweet languages	1
Users	154
User types	3
User languages	1
Date start (train)	2011-12-02 T00:47:55
Date end (train)	2012-04-10 T23:40:36
Date start (test)	2011-12-02 T00:03:32
Date end (test)	2012-04-10 T23:47:55

Table 1: Corpus statistics

Users were journalists (*periodistas*), politicians (*políticos)* or celebrities (*famosos*). The only language involved this year was Spanish (*es)*.

The list of topics that have been selected is shown in Table 2.

Topic
Politics *(política)*
Other *(otros)*
Entertainment *(entretenimiento)*
Economy *(economía)*
Music *(música)*
Soccer *(fútbol)*
Films *(cine)*
Technology *(tecnología)*
Sports *(deportes)*
Literature *(literatura)*

Table 2: Topic list

The corpus is encoded in XML (the XSD schema is provided for validation). Figure 1 shows the information of two sample tweets. The first tweet is only tagged with the global polarity (P+) and the agreement level (AGREEMENT), as the text contains no mentions to any entity, but the second one is tagged with both the global polarity of the message (P), the agreement level (AGREEMENT) and the polarity associated to each of the entities that appear in the text (UPyD and Foro Asturias, both tagged as P).


```xml
<tweet>
    <tweetid>0000000000</tweetid>
    <user>usuario0</user>
    <content>
        <![CDATA['Conozco a alguien q es adicto al drama! Ja ja ja te suena d algo!]]>
    </content>
    <date>2011-12-02T02:59:03</date>
    <lang>es</lang>
    <sentiments>
        <polarity>
            <value>P+</value>
            <type>AGREEMENT</type>
        </polarity>
    </sentiments>
    <topics>
        <topic>entretenimiento</topic>
    </topics>
</tweet>
<tweet>
    <tweetid>0000000001</tweetid>
    <user>usuario1</user>
    <content>
        <![CDATA['UPyD contará casi seguro con grupo gracias al Foro Asturias.]]>
    </content>
    <date>2011-12-02T00:21:01</date>
    <lang>es</lang>
    <sentiments>
        <polarity>
            <value>P</value>
            <type>AGREEMENT</type>
        </polarity>
        <polarity>
            <entity>UPyD</entity>
            <value>P</value>
            <type>AGREEMENT</type>
        </polarity>
        <polarity>
            <entity>Foro_Asturias</entity>
            <value>P</value>
            <type>AGREEMENT</type>
        </polarity>
    </sentiments>
    <topics>
        <topic>política</topic>
    </topics>
</tweet>
```


Figure 1: Sample tweets (General corpus)

2.2 Politics corpus

The Politics corpus contains 2 500 tweets, gathered during the electoral campaign of the 2011 general elections in Spain (Elecciones a Cortes Generales de 2011), from Twitter messages mentioning any of the four main national-level political parties: Partido Popular (PP), Partido Socialista Obrero Español (PSOE), Izquierda Unida (IU) y Unión, Progreso y Democracia (UPyD). This corpus was completely built and revised by Eugenio Martínez-Cámara, SINAI group at Universidad de Jaén, member of the organization of the task.

Similarly to the General corpus, the global polarity and the polarity at entity level for those four entities has been manually tagged for all messages. However, in this case, only 3 levels are used in this case: positive (P), neutral (NEU), negative (N), and one additional no sentiment tag (NONE).

The format is the same as the General corpus: XML as defined by the same XSD schema, where the text of the content entity has been removed to follow the Twitter restrictions. The only difference is that the entity element includes a source attribute that indicates the political party to which the entity refers: PP, PSOE, IU and UPyD.

The following figure shows the information of one sample tweet. The global polarity is N with AGREEMENT, and the polarity at entity level for the entity *@marianorajoy* whose source is *PP* is also N with AGREEMENT.


```xml
<tweet>
    <tweetid>137231808990412800</tweetid>
    <user>marianarajay</user>
    <content>
        <![CDATA['@marianorajoy Por favor, amigosh, no me votéish
          que me lo he penshado mejor con este tshunami que se me viene encima.]]>
    </content>
    <date>2011-10-17T19:13:07</date>
    <lang>es</lang>
    <sentiments>
        <polarity>
            <value>N</value>
            <type>AGREEMENT</type>
        </polarity>
        <polarity>
            <entity source="PP">@marianorajoy</entity>
            <value>N</value>
            <type>AGREEMENT</type>
        </polarity>
    </sentiments>
    <topics>
        <topic>política</topic>
    </topics>
</tweet>
```


Figure 2: Sample tweet (Politicscorpus)

The corpus will be made freely available to the community after the workshop. Please send an email to tass@daedalus.es with your email, affiliation (institution, company or any kind of organization) and a brief description of your research objectives, and you will be given a password to download the files in the password protected area. The only requirement is to include a citation to the paper (Villena-Román et al., 2013) and/or the TASS website.

3 *Description of tasks*

This year four tasks were proposed for the participants, extending the two tasks that were offered in TASS 2012. All tasks covered different aspects of sentiment analysis and automatic text classification. Registered groups could participate in one or several tasks.

Along with the submission of experiments, participants were invited to submit a paper to the workshop in order to describe their experiments and discussing the results with the audience in a regular workshop session.

These papers should follow the usual SEPLN template (as given in the author guidelines page). Reports can be written in Spanish or English. All papers were reviewed by the program committee and are included in the proceedings of the workshop and summarized in the next sections.

The four proposed tasks are described next.

3.1 Task 1: Sentiment Analysis at Global Level

This task consists on performing an automatic sentiment analysis to determine the global polarity (using 5 levels) of each message in the test set of the General corpus**.**

Participants were provided with the training set of the General corpus so that they could train and validate their models.

The evaluation metrics to evaluate and compare the different systems are the usual measurements of precision (1), recall (2) and F-measure (3) calculated over the full test set, as shown in Figure 3.

$$Precision = \frac{N\,(Correct\ classifications)}{N\,(all\ classifications)} \quad (1)$$

$$Recall = \frac{N\,(retrieved\ documents)}{N\,(all\ documents)} \quad (2)$$

$$F = (1+\beta^2)\frac{precision \cdot recall}{\beta^2 \cdot precision + recall} \quad (3)$$

Figure 3: Evaluation metrics

3.2 Task 2: Topic Classification

The technological challenge of this task is to build a classifier to automatically identify the topic of each message in the test set of the General corpus.

Participants could use the training set of the General corpus to train and validate their models.

The evaluation metrics are the same as in Task 1 (Figure 3).

3.3 Task 3: Sentiment Analysis at Entity Level

This task consists on performing an automatic sentiment analysis, similar to Task 1, but determining the polarity at entity level (using 3 polarity levels) of each message in the Politics corpus.

In this case, the polarity at entity level included in the training set of the General corpus may be used by participants to train and validate the models (converting from 5 polarity levels to 3 levels).

3.4 Task 4: Political Tendency Identification

This task moves one step forward towards reputation analysis and the objective is to estimate the political tendency of each user in the test set of the General corpus, in four possible values: LEFT, RIGHT, CENTRE and UNDEFINED.

Participants could use whatever strategy they decide, but a first approach could be to aggregate the results of the previous tasks by author and topic.

4 Participants

Participants were expected to submit one or several results of different experiments for one or several of these tasks, in the appropriate format.

Results for all tasks should be submitted in a plain text file with the following format:

```
id \t output \t confidence
```

where:

- id is the tweet ID for Tasks 1 and 2, the combination of tweet ID and entity for Task 3 (such as *142378325086715906-UPyD*), and the user ID for Task 4.
- output refers to the expected output of each task (polarity values, topic or political tendency).
- confidence is a number ranging [0, 1] that indicates the confidence in the value as assigned by the system. This value is not currently used in the evaluation.

Regarding the polarity values, there are 6 valid tags (P+, P, NEU, N, N+ and NONE). Although the polarity level must be classified into those levels and results will be evaluated for the 5 of them, the evaluation will include metrics that consider just 3 levels (POSITIVE, NEUTRAL and NEGATIVE).

Regarding the topic classification, a given tweet ID can be repeated in different lines if it is assigned more than one topic.

31 groups registered (15 groups last year) and 14 groups (9 last year) sent their submissions. The list of active participant groups is shown in Table 3, including the tasks in which they have participated.

Group	1	2	3	4
CITIUS-Cilenis	X		X	
DLSI-UA	X			
Elhuyar Fundazioa	X			
ETH-Zurich	X	X	X	X
FHC25-IMDEA		X		
ITA	X			
JRC	X			
LYS	X	X		X
SINAI-EMML	X			
SINAI-CESA	X	X	X	X
Tecnalia-UNED	X			
UNED-JRM	X	X		
UNED-LSI	X	X		
UPV	X	X	X	X
Total groups	13	7	4	4

Table 3: Participant groups

Next sections briefly describe the approaches followed by the different groups.

4.1 CITIUS-Cilenis

The joint group from Centro de Investigação em Tecnologias da Língua (CITIUS) of Universidad de Santiago de Compostela and Cilenis S.L. company sent 2 runs for Task 1 and 1 run for Task 3. Their paper (Gamallo *et al.*, 2013) describes the strategy underlying the system presented by their team for the sentiment analysis tasks. Their system is mainly based on a naive-bayes classifier for detecting the polarity of Spanish tweets. Their experiments have shown that the best performance is achieved by using a binary classifier distinguishing between just two sharp polarity categories: positive and negative.

To identify more polarity levels, the system is provided with experimentally set thresholds for detecting strong, average, and weak (or neutral) values. In addition, in order to detect tweets with and without polarity, their system makes use of a very basic rule that searches for polarity words within the analyzed text.

As it will presented in the results, this strategy is well suited to deal with coarse granularity polarity detection. Evaluation results show a good performance of the system (about 67% accuracy) when it is used to detect four sentiment categories, and its performance is significantly better when dealing with 4 (instead of 6) classification levels

Their system improves 11 points in the test evaluation, from 55% with 6 levels to 66% accuracy with 4 levels, while the improvement average of the six best systems at the TASS competition is merely 8 points.

4.2 DLSI-UA

The team from Departamento de Lenguajes y Sistemas Informáticos at Universidad de Alicante submitted 3 runs for Task 1.

Their contribution (Fernández *et al.*, 2013) consists of two different approaches: a modified version of a ranking algorithm (RA-SR) using bigrams, used on the Task 2 of the Semeval 2013 competition, and a new proposal using a skipgrams scorer. Both approaches create sentiment lexicons able to retain the context of the terms, and employ machine learning techniques to detect the polarity of a text.

To create the sentiment resource they used the training set of the General corpus, and a classifier is created using the Weka default implementation of the Support Vector Machines (SVM) algorithm using the features described in their paper.

All their approaches appear in the top 10 best results of the systems presented to the competition, and the combination of them reaches the first position.

4.3 Elhuyar Fundazioa

(Saralegi and San Vicente, 2013) describes the system presented for Task 1 of sentiment analysis, where 2 runs where submitted. They adopted a supervised approach using a SVM classifier that includes some linguistic knowledge-based processing for preparing the features.

Their system effectively combines several features based on linguistic knowledge. The processing comprises lemmatization, POS tagging, tagging of polarity words, treatment of emoticons and treatment of negation. A pre-processing for treatment of spell-errors is also performed.

Detection of polarity words is done according to a polarity lexicon built in two ways: projection to Spanish of an English lexicon, an extraction of divergent words of positive and negative tweets of training corpus.

In their case, using a semi-automatically built polarity lexicon improves the system performance significantly over a unigram model. Other features such as POS tags, and especially word polarity statistics were also found to be helpful. They improved the tweet normalization step over last year's algorithm.

Their system achieves a 60% accuracy for fine granularity and a 68% accuracy for coarse granularity polarity detection. Overall, the system shows robust performance when it is evaluated against test data different from the training data.

4.4 ETH-Zurich

(García and Thelwall, 2013) describe the participation of ETH-Zurich team in the four tasks. They present a study political discourse and emotional expression through a dataset of Spanish tweets. They analyze the political position of four major parties through their Twitter activity, revealing that Twitter political discourse depends on subjective perception, and resembles the political space of Spain.

They propose a simplified lexicon-based method to identify the topics of a tweet, which works especially well to detect the political content of tweets.

Furthermore, they adapted SentiStrength to Spanish, by translating and converting an established lexicon of word valence. Under certain design decisions, this tool performs better than random, with ample room for improvement. Finally, they combined three datasets to analyze the sentiment expressed in the political tweets of four major Spanish parties, finding differences related to the status quo, and the Spanish political climate.

Their manual crowdsource approach to political tendency detection (Task 4) achieves a 73% precision of the test corpus created by pooling.

4.5 FHC25-IMDEA

(Cordobés *et al.*, 2013) describe the participation of this combined team formed by Institute IMDEA Networks, U-tad and Factor Holding Company 25 companies.

This group has submitted 2 runs for Task 2 regarding topic classification of texts, using a technique based on graph similarity to classify Twitter messages as being related to a specific topic.

Their core approach is that any text can be represented as a graph. For a given text, their system places the terms (actually the stems) in the vertexes of a graph and creates links with a given weight among them. Then their hypothesis is that graphs belonging to texts of the same topic usually form unique structures (i.e., a topic graph). Thus, they use a metric for calculating the similarity between the text graph to classify and the different topic graphs.

Their system achieves a 71,9% precision value. The analysis per category shows that the system is very biased towards the most frequent topics (politics and others), which cover 46,4% and 49,5% of the total number of tweets, and achieve about a 78% in precision value. The rest of the topics show a poorer performance, many of them below 50%.

4.6 ITA

This group from Instituto Tecnológico de Aragón (Del Hoyo *et al.*, 2013) made some experiments for Task 1 with the Non-Axiomatic Reasoning System (NARS) as a tool to dynamically discover content words and phrases with opinion.

New techniques based in Artificial General Intelligence such as NARS, aims to explain a large variety of cognitive phenomena with a unified theory. What makes NARS different from conventional reasoning systems is its ability to learn from its experience and to work with insufficient knowledge and resources, from a logic perspective. NARS attempts to uniformly explain and reproduce many cognitive facilities.

They intend to use its logic and reasoning capabilities for modeling human language and thereby identifying the polarity of the new words. The main idea of their experiment is to use a seed dictionary to look for new similar polarity words.

4.7 JRC

(Balahur and Perea-Ortega, 2013) from European Commission, Joint Research Centre (JRC) presents several experiments for the task entitled sentiment analysis at global level (Task 1) within the TASS 2013 evaluation campaign.

To tackle this task, an approach based on machine learning by trying different feature combinations was applied. Several in-house built dictionaries and machine translated data for training were employed by adapting an approach designed for English to Spanish.

Additionally, four separate classifiers were tested in cascade to determine the sentiment from the general to the finer-grained classes of polarity.

Although this was their first participation, the proposed approaches might be considered good strategies to generate learning data for polarity classification systems in Spanish.

4.8 LYS

The paper from the team at Departamento de Computación, Universidade da Coruña (Vilares *et al.*, 2013) describes the approach developed by their group in order to solve the sentiment analysis at a global level (Task 1), topic identification (Task 2) and political tendency classification tasks on Spanish tweets (Task 4).

As a preliminary step, they carry out an ad-hoc preprocessing in order to normalize the tweets. They then apply part-of-speech tagging and dependency parsing algorithms to the tweets to obtain their syntactic structure.

Their proposal also employs psychological resources in order to exploit the psychometric properties of human language.

The experimental results confirm the robustness of the proposal, which has achieved good performance in general, being the best-performing approach in the topic classification task.

4.9 SINAI-EMML

(Martínez Cámara *et al.*, 2013) present the participation of the SINAI-EMML research group of the University of Jaén in Task 1. Their participation has focused on the polarity classification tweets in Spanish, with 4 and 6 classes.

They opted for a completely unsupervised strategy in order to get results and conclusions that help them to improve our supervised system developed and tested in TASS 2012.

This system is based on the combination of three linguistic resources, SentiWordNet, Q-WordNet and iSOL.

The calculation of the polarity value has been made based on the addition of the differences between the positive and negative values of each term, normalized with the max value.

4.10 SINAI-CESA

(Montejo-Ráez *et al.*, 2013) describes the participation of the CESA team of the SINAI group in the four proposed tasks. Their system proposes a solution based on Information Retrieval, by applying Latent Semantic Analysis (LSA).

Their approach takes its train data from the continuous stream of posts from Twitter, capturing those that are likely to include affective expressions and generating a corpus of “feelings” that are labeled according to their polarity. No training data from controlled corpora have been used, as they believe that trained models suffer from domain related limitations.

Results are not very promising compared to other competitors, but the method opens a new approach in the use of social web publications as resource for sentiment polarity classification.

4.11 Tecnalia-UNED

The participation of the joint Tecnalia-UNED team with 1 run in Task 1 is fully described in (Villar Rodríguez *et al.*, 2013). Based on a previous analysis of the TASS 2012 corpus, their system executes a stage by stage advanced linguistic process, further than the basic machine learning approach adopted by many research groups, trying to deal with complex issues such as negation detection and emphatiser treatment (aiming at distinguishing the range of polarity levels). The linguistic processing relies on the Freeling tool.

Their results, though not ranked at top participants, are good enough to draw some conclusions. The modifier processing module is one of the bottlenecks as it seems not to be adapted to the use of emphatisers in the corpus, or the existence of other semantic phenomena.

In addition, the overall calculation of polarity, based on the aggregation of the polarity of different segments or fragments, is not the best choice as values at the extremes of the range (P+ and N+) seem to be reduced.

4.12 UNED-JRM

(Rufo Mendo, 2013) at UNED aims to give a new vision to the work of topic classification of text and sentiment analysis for Task 1 and 2. Although both of them are classification tasks, they are usually addressed differently. In their work, they carried out the development of a classifier that performs the two tasks indifferently, with similar results and checking that perhaps there can be a single solution for classification tasks.

It also presents an analysis of the behavior of a supervised classifier compared to semi-supervised classifier.

Results have not been very encouraging, their experiments ranked 30th out of 36 in Task 1 with 5 levels, position 21 of 46 in Task 1 with 3 levels, and position 7 of 20 in Task 2. This encourages to further research in this area, trying to reach new editions of TASS with better results and larger contributions.

4.13 UNED-LSI

(Castellanos González *et al.*, 2013) summarizes the work proposed for the participation of LSI group at UNED our participation in Task 1 and Task 2. Their contribution is an extension of a previous work done for TASS 2012.

The work carried out in the previous year was focused on the tweet classification based on an Information Retrieval (IR) approach: the classes are modeled according to the textual information of the tweets belonging to each class, and the tweets classify are used as query.

This year they have applied this approach on sentiment analysis and topic classification tasks, but their work has focused on analyzing the type of tweet information to use to carry out the classification and what process should be followed to take this information into account.

In this sense, they have proposed different types of modeling as well as different ways of performing the information retrieval process according to the different types of information.

The results suggest that although the use of this type of information is valuable (especially named entities), it should always be done in conjunction with the overall tweet contents.

4.14 UPV

(Pla and Hurtado, 2013) describes the participation of the ELiRF research group of the Universitat Politècnica de València in all tasks, with good results in all them. This work describes the approaches used, the results obtained and a discussion of these results.

They state that the first basic process is to carry out an adequate tokenization of the tweets, and this should be done with specific adaptations to any of the different tools that are publicly available, due to the type of language used in social networks (non-grammatical phrases, lack or misuse of punctuation symbols, specific terminology, slang, etc.).

Moreover, using other basic resources in Natural Language Processing, all of them based in normative text, is unfeasible without corrective actions. In their experiments, they have used and adapted to Spanish the tweet tokenizer called Tweetmotif and also Freeling.

For the classification stage, Support Vector Machines have been used, my means of Weka and the external LibSVM library.

Results achieved are similar, and some times, even the top ranked in the competition.

5 *Results*

After the submission deadline, runs were collected and checked and results were evaluated by the organization and made available to the participants (downloaded from the password protected are in the website) to allow them to prepare their reports.

Results included a spreadsheet with the overall values for each task, including precision, recall and F-Measure, and also detailed results per experiment for all the 5 evaluations (as explained before, Task 1 was evaluated using both 5-level and 3-level setups), the confusion matrix with all labels to allow error analysis, and finally the gold standard for the task (or qrels in TREC context) itself. The PHP script used for the evaluation of each submission was also included for their convenience.

Results for each task are described next.

5.1 Task 1: Sentiment Analysis at Global Level

The number of runs submitted to this task is shown in Table 4.

Group	Runs
CITIUS-Cilenis	2
DLSI-UA	3
Elhuyar Fundazioa	2
ETH-Zurich	3
ITA	1
JRC	18
LYS	2
SINAI-EMML	2
SINAI-CESA	2
Tecnalia-UNED	1
UNED-JRM	2
UNED-LSI	15
UPV	3
Total runs	56*

** 10 submissions specific for 3 levels*

Table 4: Number of runs for Task 1

Initially a gold standard was generated by pooling all submissions with a votation schema and then an extensive human review of the ambiguous decisions was carried out. However, as the pooling process was made by using all runs and some groups had submitted many runs and other groups had only submitted a few, some concern arose about a possible bias for those groups.

To avoid any systemic problem, the gold standard creation should be repeated or at least carefully evaluated for correctness. Due to the summer holidays and lack of resources, finally the gold standard of TASS 2012 has to be used to evaluate the submissions. That gold standard was not subject to this bias as the number of submissions was balanced in the previous edition.

Results are listed in the tables below. All tables show the precision (P), recall (R) and F1 value achieved in each experiment. Table 5 considers 5 levels of sentiments (P+, P, NEU, N, N+) and no sentiment (NONE).

Precision values range from 61.6% to 12.6%. The average values are 43.3% for all metrics.

Run Id	P	R	F1
DLSI-UA-pol-dlsiua3-3-5l.txt	0.616	0.616	0.616
Elhuyar-TASS2013_Elhuyar_run1	0.601	0.601	0.601
Elhuyar-TASS2013_Elhuyar_run2	0.599	0.599	0.599
DLSI-UA-pol-dlsiua3-2-5l.txt	0.596	0.596	0.596
UPV_ELiRF_task1_run2.txt	0.576	0.576	0.576
UPV_ELiRF_task1_run1.txt	0.574	0.574	0.574
UPV_ELiRF_task1_run3.txt	0.573	0.573	0.573
CITIUS-task1_CITIUS_1.txt	0.558	0.558	0.558
lys_global_sentiment_task_6c	0.553	0.553	0.553
DLSI-UA-pol-dlsiua3-1-5l.txt	0.552	0.552	0.552
CITIUS-task1_CITIUS_2.txt	0.541	0.541	0.541
lys_global_sentiment_task_6c_wui	0.533	0.533	0.533
JRC-tassTrain-base-DICT-5way.tsv	0.519	0.519	0.519
JRC-tassTrain-lemmaStop-SVM-5way.tsv	0.515	0.515	0.515
JRC-tassTrain-lemmaStop-DICT-5way.tsv	0.507	0.507	0.507
JRC-tassTrain-base-SVM-5way.tsv	0.505	0.505	0.505
JRC-tassTrain-lemma-SVM-5way.tsv	0.504	0.504	0.504
JRC-tassTrain-lemma-DICT-5way.tsv	0.497	0.497	0.497
JRC-tassTrain-lemmaStop-4CLS-5way.tsv	0.481	0.481	0.481
JRC-tassTrain-base-4CLS-5way.tsv	0.477	0.477	0.477
JRC-tassTrain-lemma-4CLS-5way.tsv	0.477	0.477	0.477
ITA_ResultadosAnalisisOpiniónAlg	0.439	0.439	0.439
LSI_UNED_2_TASK1_RUN_09	0.402	0.402	0.402
LSI_UNED_2_TASK1_RUN_14	0.402	0.402	0.402
LSI_UNED_2_TASK1_RUN_04	0.398	0.398	0.398
LSI_UNED_2_TASK1_RUN_11	0.398	0.398	0.398
LSI_UNED_2_TASK1_RUN_15	0.396	0.396	0.396
LSI_UNED_2_TASK1_RUN_05	0.395	0.395	0.395
LSI_UNED_2_TASK1_RUN_07	0.395	0.395	0.395
LSI_UNED_2_TASK1_RUN_10	0.395	0.395	0.395
LSI_UNED_2_TASK1_RUN_02	0.393	0.393	0.393
UNED-JRM-task1-run2.txt	0.393	0.393	0.393
LSI_UNED_2_TASK1_RUN_03	0.391	0.391	0.391
LSI_UNED_2_TASK1_RUN_08	0.391	0.391	0.391
LSI_UNED_2_TASK1_RUN_12	0.386	0.386	0.386
LSI_UNED_2_TASK1_RUN_13	0.386	0.386	0.386
LSI_UNED_2_TASK1_RUN_06	0.359	0.359	0.359
LSI_UNED_2_TASK1_RUN_01	0.354	0.354	0.354
TECNALIA-UNED.txt	0.348	0.344	0.346
ETH-task1-Warriner.txt	0.328	0.328	0.328
sinai_emml_task1_6classes.txt	0.314	0.314	0.314
ETH-task1-OptT1.txt	0.249	0.249	0.249
ETH-task1-OptT2.txt	0.244	0.244	0.244
sinai_cesa-task1_raw.tsv	0.135	0.134	0.134
sinai_cesa-task1_normalized.tsv	0.131	0.131	0.131
UNED-JRM-task1.txt	0.126	0.126	0.126

Table 5: Results for task 1 (Sentiment Analysis) with 5 levels + NONE

In order to perform a more in-depth evaluation, Table 6 gives results considering the classification only in 3 levels (POS, NEU, NEG) and no sentiment (NONE) merging P and P+ in only one category, as well as N and N+ in another one.

In this case, precision values improve, as expected. The precision obtained now ranges from 68.6% to 23.0%. The average values for all metrics in this case is 53.0%.

Run Id	P	R	F1
Elhuyar-TASS2013_Elhuyar_run1	0.686	0.686	0.686
Elhuyar-TASS2013_Elhuyar_run2	0.684	0.684	0.684
UPV_ELiRF_task1_run2.txt	0.674	0.674	0.674
UPV_ELiRF_task1_run3.txt	0.674	0.674	0.674
UPV_ELiRF_task1_run1.txt	0.672	0.672	0.672
CITIUS-task1_CITIUS_1.txt	0.668	0.668	0.668
DLSI-UA-pol-dlsiua3-3-5l.txt	0.663	0.663	0.663

lys_global_sentiment_task_6c	0.657	0.657	0.657
lys_global_sentiment_task_6c_wui	0.647	0.647	0.647
DLSI-UA-pol-dlsiua3-2-5l.txt	0.640	0.640	0.640
CITIUS-task1_CITIUS_2.txt	0.622	0.622	0.622
DLSI-UA-pol-dlsiua3-1-5l.txt	0.620	0.620	0.620
JRC-tassTrain-base-DICT-3way.tsv	0.612	0.612	0.612
JRC-tassTrain-lemmaStop-SVM-3way.tsv	0.608	0.608	0.608
JRC-tassTrain-lemmaStop-DICT-3way.tsv	0.607	0.607	0.607
JRC-tassTrain-lemma-DICT-3way.tsv	0.599	0.599	0.599
JRC-tassTrain-lemma-SVM-3way.tsv	0.599	0.599	0.599
JRC-tassTrain-base-SVM-3way.tsv	0.597	0.597	0.597
JRC-semevaltassTrain-base-DICT-3way.tsv	0.590	0.590	0.590
JRC-semevaltassTrain-base-SVM-3way.tsv	0.585	0.585	0.585
JRC-tassTrain-lemmaStop-4CLS-3way.tsv	0.582	0.582	0.582
ITA_ResultadosAnalisisOpiniónAlg	0.543	0.543	0.543
TECNALIA-UNED.txt	0.496	0.490	0.493
UNED-JRM-task1-run2.txt	0.496	0.496	0.496
LSI_UNED_2_TASK1_RUN_06	0.479	0.479	0.479
LSI_UNED_2_TASK1_RUN_07	0.476	0.476	0.476
LSI_UNED_2_TASK1_RUN_02	0.474	0.474	0.474
LSI_UNED_2_TASK1_RUN_01	0.471	0.471	0.471
LSI_UNED_2_TASK1_RUN_08	0.470	0.470	0.470
LSI_UNED_2_TASK1_RUN_03	0.467	0.467	0.467
ETH-task1-Warriner.txt	0.466	0.466	0.466
LSI_UNED_2_TASK1_RUN_09	0.464	0.464	0.464
ETH-task1-OptT2.txt	0.461	0.461	0.461
LSI_UNED_2_TASK1_RUN_04	0.461	0.461	0.461
LSI_UNED_2_TASK1_RUN_11	0.459	0.459	0.459
LSI_UNED_2_TASK1_RUN_10	0.457	0.457	0.457
LSI_UNED_2_TASK1_RUN_05	0.454	0.454	0.454
ETH-task1-OptT1.txt	0.441	0.441	0.441
sinai_emml_task1_3classes.txt	0.409	0.409	0.409
LSI_UNED_2_TASK1_RUN_14	0.408	0.408	0.408
LSI_UNED_2_TASK1_RUN_15	0.408	0.408	0.408
LSI_UNED_2_TASK1_RUN_13	0.407	0.407	0.407
LSI_UNED_2_TASK1_RUN_12	0.405	0.405	0.405
sinai_cesa-task1_raw.tsv	0.389	0.388	0.388
sinai_cesa-task1_normalized.tsv	0.388	0.387	0.387
UNED-JRM-task1.txt	0.230	0.230	0.230

Table 6: Results task 1 (Sentiment Analysis) with 3 levels + NONE

The distribution of labels in both the training and test corpus is shown in Table 7. Obviously, the distribution among different sentiment labels is not evenly balanced in both corpus, i.e., the gold standard may be not well built. This fact causes that, for example, given a system that is able to correctly classify P+ and NONE with a high precision (both of them count for the 70% of the tweets in the test corpus), and maybe, not so good at classifying the other labels, may achieve better results on the test corpus than the training corpus, as it is actually reported by some participants (CITIUS-Cilenis and Elhuyar).

Sentiment	Frequency (Train)	Frequency (Test)
P+	22.44%	34.12%
P	4.12%	2.45%
NEU	8.45%	2.15%
N	16.91%	18.56%
N+	12.51%	7.5%
NONE	23.58%	35.22%

Table 7: Sentiment distribution

This is for example the case of *CITIUS-task1_CITIUS_1.txt* run, which achieves better results than *lys_global_sentiment_task_6c*, but is worse balanced.

Sentiment	Precision CITIUS	Precision LYS
P+	0.791	0.578
P	0.363	0.569
NEU	0.022	0.195
N	0.289	0.548
N+	0.533	0.526
NONE	0.546	0.557

Table 8: Comparison of precision per sentiment label (CITIUS vs LYS)

Another interesting comparison is the top ranked run, *DLSI-UA-pol-dlsiua3-3-5l.txt*, vs the second ranked, *TASS2013_Elhuyar_run1*, shown in Table 9. Results from Elhuyar are quite balanced and can be compared to the LYS run, but they are better ranked as they achieve greater precision for all labels but N+ and NEU. In turn, results from DLSI are better than Elhuyar run because their system performs better for P+ and NONE that are the most frequent labels. This issue must be studied for eventual future editions of TASS.

Sentiment	Precision DLSI	Precision Elhuyar
P+	0.705	0.638
P	0.263	0.661
NEU	0.108	0.185
N	0.586	0.583
N+	0.390	0.427
NONE	0.649	0.631

Table 9: Comparison of precision per sentiment label (DLSI vs Elhuyar)

5.2 Task 2: Topic Classification

This task has been evaluated as a single label classification. The most restrictive criterion has been applied: a "success" is achieved only when all the test labels have been returned. Participants were welcomed to reevaluate their experiments with a less restrictive strategy in their papers.

Similarly to the first task, the gold standard finally considered was the one used in TASS 2012. The distribution of topics in both the training and test corpus is shown in Table 10, sorted by frequency in the train corpus. The total frequency is greater than the number of tweets as several topics can be assigned per tweet.

Topic	Frequency (Train)	Frequency (Test)
Politics	3 120 (33%)	30 067 (43%)
Other	2 337 (24%)	28 191 (40%)
Entertainment	1 678 (17%)	5 421 (8%)
Economy	942 (10%)	2 549 (3%)
Music	566 (6%)	1 498 (2%)
Soccer	252 (3%)	823 (1%)
Films	245 (3%)	596 (1%)
Technology	217 (2%)	287 (0%)
Sports	113 (1%)	135(0%)
Literature	103 (1%)	93(0%)
TOTAL	9 573	69 660

Table 10: Topic distribution

Table 11 shows the results for Task 2. 20 experiments were submitted in all. In this task, precision ranges from 80.4% to 16.1%. The average values are 62.4% precision, 44.4% recall and 49.6 F1. As in task 1, different submissions from the same group usually have a similar values.

Run Id	P	R	F1
lys_topic_task_with_user_info.qrel	0.804	0.804	0.804
lys_topic_task.qrel	0.786	0.786	0.786
LSI_UNED_2_TASK2_RUN_07	0.777	0.184	0.298
LSI_UNED_2_TASK2_RUN_08	0.773	0.158	0.262
UPV_ELiRF_task2_run2.txt	0.756	0.756	0.756
UPV_ELiRF_task2_run1.txt	0.755	0.755	0.755
ETH-task2.txt	0.734	0.455	0.562
LSI_UNED_2_TASK2_RUN_09	0.727	0.211	0.327
FHC25-IMDEAults_PR_GD_TT.txt.res	0.719	0.702	0.710
FHC25-IMDEAults_PR_TT.txt.res	0.705	0.688	0.696
LSI_UNED_2_TASK2_RUN_01	0.660	0.404	0.501
LSI_UNED_2_TASK2_RUN_04	0.659	0.400	0.498
LSI_UNED_2_TASK2_RUN_03	0.653	0.406	0.501
LSI_UNED_2_TASK2_RUN_02	0.649	0.381	0.480
LSI_UNED_2_TASK2_RUN_06	0.646	0.377	0.476
LSI_UNED_2_TASK2_RUN_05	0.639	0.366	0.465
UNED-JRM-task2-run2.txt	0.479	0.479	0.479
UNED-JRM-task2.txt	0.240	0.240	0.240
sinai_cesa-task2_normalized.tsv	0.161	0.159	0.160
sinai_cesa-task2_raw.tsv	0.161	0.159	0.160

Table 11: Results for task 2 (Topic classification)

Some participants pointed out (such as FHC25-IMDEA) that, as shown in Table 10, the distribution is quite balanced between both corpus but not on different topics. This may cause that the trained systems tend to be biased towards the most frequent topics (politics and other). Systems that are optimized for those categories, even at the cost of a low performance in the less frequent topics, will seem to achieve a better overall result than a system that is more balanced system.

5.3 Task 3: Sentiment Analysis at Entity Level

The evaluation is made over the Politics corpus, which was tagged manually, so the gold standard is created with no pooling. Finally 6 runs were submitted for this task. Results are shown in Table 12.

Average precision is 37.2%, recall is 36.5% and F1 is 36.9%. These figures are much lower than the values achieved in Task 1. This is obviously because Task 3 is considerably harder than Task 1 and systems do not reach the adequate level of development.

Run Id	P	R	F1
CITIUS-task3_CITIUS.txt	0.411	0.378	0.394
UPV_ELiRF_task3_run0.txt	0.395	0.395	0.395
sinai_cesa-task3_normalized.tsv	0.384	0.384	0.384
sinai_cesa-task3_raw.tsv	0.376	0.372	0.374
UPV_ELiRF_task3_run1.txt	0.358	0.357	0.357
ETH-task3.txt	0.307	0.307	0.307

Table 12: Results for task 3 (Sentiment Analysis at Entity Level)

5.4 Task 4: Political Tendency Identification

The gold standard was built manually by reviewing each user's political tendency, as defined by himself/herself, or assigning UNDEFINED if not stated or unknown.

Finally 11 runs were submitted. Results are shown in Table 13.

Run Id	P	R	F1
ETH-Task4-Crowdsource.txt [MANUAL]	0.734	0.734	0.734
UPV_ELiRF_task4_run1.txt	0.703	0.703	0.703
UPV_ELiRF_task4_run4.txt	0.696	0.696	0.696
UPV_ELiRF_task4_run2.txt	0.677	0.677	0.677
UPV_ELiRF_task4_run3.txt	0.658	0.658	0.658
sinai_cesa-task4_nound_raw.tsv	0.583	0.399	0.474
sinai_cesa-task4_nound_normalized.tsv	0.570	0.386	0.460
sinai_cesa-task4_und_normalized.tsv	0.467	0.316	0.377
sinai_cesa-task4_und_raw.tsv	0.444	0.304	0.361
lys_political_tendency_task_model2	0.424	0.424	0.424
lys_political_tendency_task_model1	0.386	0.386	0.386

Table 13: Results for task 4 (Political Tendency Identification)

Average values for precision, recall and F1 are 57.7%, 51.7% and 54.1% respectively. Run from ETH is based on a manual assignment of political tendency to each user, made with crowdsourcing, so it is supposed to achieve the best result in the gold standard, as it happens.

6 *Conclusions and Future Work*

TASS was the first workshop about sentiment analysis in the context of SEPLN. This second edition of TASS has been even more successful than the first one, as the number of participants has increased up to 31 groups registered (15 groups last year) and 14 groups (9 last year) sent their submissions. We think that the number of participants, the quality of their work and their reports, and the good results achieved in such hard tasks, has met and gone beyond all our expectations.

It is still necessary to perform a more detailed analysis of the results. However, the developed corpus and gold standards, and the reports from participants will for sure be helpful for other research groups approaching these tasks.

TASS 2012 corpus, released after the workshop for free use by the research community, has been downloaded up to date by more than 50 research groups, 17 out of Spain. We expect to reach a similar impact with TASS 2013 corpus.

Acknowledgements

This work has been partially supported by several Spanish research projects: Ciudad2020: Hacia un nuevo modelo de ciudad inteligente sostenible (INNPRONTA, CDTI, IPT-20111006, 2011-2014), MA2VICMR: Improving the access, analysis and visibility of the multilingual and multimedia information in web for the Region of Madrid (S2009/TIC-1542) and MULTIMEDICA: Multilingual Information Extraction in Health domain and application to scientific and informative documents (TIN2010-20644-C03-01). Authors would like to thank all partners for their knowledge and support.

References

Villena-Román, Julio; Lana-Serrano, Sara; Martínez-Cámara, Eugenio; González-Cristobal, José Carlos. 2013. TASS - Workshop on Sentiment Analysis at SEPLN. Revista de Procesamiento del Lenguaje Natural, 50, pp 37-44. http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/4657.

Gamallo, Pablo; García, Marcos; Fernández-Lanza, Santiago. 2013. *TASS: A Naive-Bayes strategy for sentiment analysis on Spanish tweets.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

Fernández, Javi; Gutiérrez, Yoan; Gómez, José M.; Martínez-Barco, Patricio; Montoyo, Andrés; Muñoz, Rafael. 2013. *Sentiment Analysis of Spanish Tweets Using a Ranking Algorithm and Skipgrams.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

Saralegi Urizar, Xabier; San Vicente Roncal, Iñaki. 2013. *Elhuyar at TASS 2013.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

García, David; Thelwall, Mike. 2013. *Political alignment and emotional expression in Spanish Tweets.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

Cordobés, Héctor; Fernández-Anta, Antonio; Núñez, Luis Felipe; Pérez, Fernando; Redondo, Teófilo; Santos, Agustín. 2013. *Técnicas basadas en grafos para la categorización de tweets por tema.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

Del Hoyo, Rafael; Hupont, Isabelle; Lacueva, Francisco. 2013. *Affective Polarity word discovering by means of Artificial General Intelligence techniques.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

Balahur, Alexandra; Perea-Ortega, José M. 2013. *Experiments using varying sizes and machine translated data for sentiment analysis in Twitter.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

Vilares, David; Alonso, Miguel A.; Gómez-Rodríguez, Carlos. 2013. *LyS at TASS 2013: Analysing Spanish tweets by means of dependency parsing, semantic-oriented lexicons and psychometric word-properties.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

Martínez Cámara, Eugenio; García Cumbreras, Miguel Ángel; Martín Valdivia, M. Teresa; Ureña López, L. Alfonso. 2013. *SINAI-EMML: SINAI-EMML: Combinación de Recursos Lingüíticos para el Análisis de la Opinión en Twitter.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

Montejo-Ráez, A.; Díaz-Galiano, M. C.; García-Vega, M. 2013. *LSA based approach to TASS 2013.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

Villar Rodríguez, Esther; Torre Bastida, Ana Isabel; García Serrano, Ana; González Rodríguez, Marta. 2013. *TECNALIA-UNED @ TASS: Uso de un enfoque lingüístico para el análisis de sentimientos.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

Rufo Mendo, Francisco Javier. 2013. *Are topic classification and sentiment analysis really different?* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

Castellanos González, Ángel; Cigarrán Recuero, Juan; García Serrano, Ana. 2013. *UNED LSI @ TASS 2013: Considerations about Textual Representation for IR based Tweet Classification.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.

Pla, Ferran; Hurtado, Lluís-F. 2013. *ELiRF-UPV en TASS-2013: Análisis de Sentimientos en Twitter.* In Proc. of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.