Spanish Knowledge Base Generation for Polarity Classification from Masses

Arturo Montejo-Ráez*
Computer Science Department
University of Jaén
23071 Jaén
Spain
amontejo@ujaen.es

Manuel C. Díaz-Galiano
Computer Science Department
University of Jaén
23071 Jaén
Spain
mcdiaz@ujaen.es

José M. Perea-Ortega
Computer Science Department
University of Jaén
23071 Jaén
Spain
jmperea@ujaen.es

L. Alfonso Ureña-López
Computer Science Department
University of Jaén
23071 Jaén
Spain
laurena@ujaen.es

ABSTRACT

This work presents a novel method for the generation of a knowledge base oriented to Sentiment Analysis from the continuous stream of published micro-blogs in social media services like Twitter. The method is simple in its approach and has shown to be effective compared to other knowledge based methods for Polarity Classification. Due to independence from language, the method has been tested on different Spanish corpora, with a minimal effort in the lexical resources involved. Although for two of the three studied corpora the obtained results did not improve those officially obtained on the same corpora, it should be noted that this is an unsupervised approach and the accuracy levels achieved were close to those levels obtained with well-known supervised algorithms.

Categories and Subject Descriptors

H.3.1 [**Content Analysis and Indexing**]: Linguistic processing

General Terms

Social media, Knowledge-base generation, Sentiment Analysis, Polarity Classification

Keywords

ACM proceedings, LaTeX, text tagging

1. INTRODUCTION

Nowadays the World Wide Web is a vast information resource where people tend to disclose their opinions and sentiments. For example, many protests, criticisms or organized activities are planned using Facebook, Twitter and even blogs. Opinions on various topics can be expressed in unstructured documents, reviews, posts, comments, etc. Tackling and tracking this huge unstructured information in order to detect its polarity is attracting many researchers in the field of text mining.

Two main types of information can be found on the Internet: facts and opinions. Although there are lots of issues to be resolved, the management of factual information has been extensively studied. However, the automatic processing of textual opinions is a new task closely related to text mining, which has just started to be studied. This is a challenging task known as Opinion Mining (OM), sometimes also called Sentiment Analysis (SA) [25]. This new discipline aims to identify and analyze opinions and emotions. It includes several sub-tasks such as subjectivity detection, polarity classification, review summarization, humor detection or emotion classification, among others. Specifically, sentiment classification or polarity detection is an opinion mining activity oriented to determine which is the overall sentiment-orientation of the opinions contained within a given document [26]. The document is supposed to contain subjective information such as product reviews or opinionated posts in blogs.

This paper introduces a novel approach in the generation of knowledge resources for Sentiment Analysis by crawling the vast flow of micro-texts published in social media every second. By filtering an small (but yet a huge) part of these streams and categorizing them semi-automatically, we have been able to produce a resource for Polarity Classification with little human intervention. The idea behind could be summarized as "*Let the crowd help you to know about the crowd*". Thus, by crawling tweets and categorizing them in basis of simple regular expressions similar to *Me siento X* (*I feel like X* we can built automatically a big collection of data, as explained later. In this way, we categorize tweets by means of tweets, resulting in a promising method for many other problems and in a knowledge-base generation

*Referring author

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

in constant evolution thanks to the ever lasting stream of posts.

The rest of the paper is organized as follows: the next section presents work related to polarity detection dealing with languages other than English and multilingual opinion mining. Section 3 describes the procedure followed to generate the corpus proposed in this study, along with its main features. Section 4 describes the different resources used in our experiments, as well as the corpora employed for comparison purposes. Section 5 describes the polarity classification approach proposed in this study. The different experiments carried out and the results obtained are expounded in Section 6, also showing the comparison with other works. Finally, the main conclusions and ideas for further work are detailed in Section 7.

2. BACKGROUND

Although Opinion Mining (OM) is a relatively new discipline, there is a considerable number of research works on this subject. A good review on Opinion Mining and Sentiment Analysis can be found in [25]. This work describes some useful resources and tools for OM and also comments the main contributions in this field. Although different approaches have been applied in the field of sentiment-polarity classification, the mainstream basically consists of two major methodologies: supervised and unsupervised approaches. On the one hand the supervised approach is based on using a collection of data to train the classifiers [26]. On the other hand, the unsupervised approach, also known as Semantic Orientation (SO) approach, does not need prior training, but it takes into account the orientation of words, positive or negative [31]. Both methodologies have their advantages and drawbacks. For example, the supervised approach requires training data, which in many cases are impossible or difficult to achieve, partially due to the novelty of the task or even the language used. In opposition, the unsupervised approach requires having lots of linguistic resources which generally depend on the language.

Language resources for natural language processing are very valuable. The construction and generation of such resources is a hard work that may take several years and a significant investment. Knowledge bases like WordNet [13], WordNet-Affect [32] or SenticNet [6] have been found to be useful in many problems [18], although their cover range on other languages is far from its original language (English), ranging from the well-known EuroWordNet lexical database [33] to adaptations to non-covered languages, like Chinese [9] and more recent studies in other languages [21].

In Sentiment Analysis, the use of lexical databases has motivated intensive research work during past years [27, 3], with some relevant resources like SentiWordNet [2]. Again, resources on other languages and domains (informal and subjective texts) are difficult to obtain, so new ways of exploring how to represent textual streams with the knowledge implicit in the stream itself are rising [36].

The language used in social media is distant from formal language used in press and other controlled sources. People express their states in many ways, using a rich jargon and compressed forms (abbreviations), being Twitter the main example of this type of communication and broadcasting [19]. Thus, a normalization process must be performed in order to apply traditional language resources. These resources expect a proper grammar and vocabulary, so its effectiveness is seriously affected by the challenging style of writing in micro-blogs [16].

In any case, the stream of texts provided every second by Twitter must be taking as a valuable resource by itself. In fact, Twitter has been found to be very useful in many scenarios, like real-time Recommender Systems [11], cinema revenue prediction [1] or even crime prediction [34], among others. In Sentiment Analysis, Twitter has gain a prominent role [22], and nowadays is the subject of a very active research community.

2.1 Generating a resource from massive information

The generation of resources for sentiment analysis is not new, and is still of interest due to the lack of non-English knowledge bases or lexicons. Banea et al. proposed a method for generating these kind of resources from existing ones [5]. Their concerns were similar to ours, but we even try to not suposse initial "seed words" or starting set of emotional terms. Our hypothesis is that, despite the variety and uncontrolled vocabulary of tweets, the volume of the stream is big enough to consider that controlled filters on released posts can lead to a rough corpus whose utility will depend on the nature and focus of the filters applied. This hypothesis is similar to that of Google Flue Trends[1], which dives into people searches across time in order to monitor disease outbreaks, like flue [14, 12, 8]. In general, the real-time analysis of streaming texts is of growing interest to many entities, like governments and companies and some interesting and visual solutions are appearing [17]. The use of the micro-blogging web as a source for social monitoring is the goal of many projects nowadays [10], and numerous tools appear every day [29].

Following the line drawn by previous works where texts are not directly processed but projected to a different conceptual (or emotional) space [7], our approach proposes to create a vector of emotions (i.e. a emotional indexing of tweets) by performing a search with the tweet as a query over a self-generated corpus.

3. MESIENTO: AN EVOLVING SPANISH CORPUS FOR SENTIMENT INDEXING

In this section we present the *MeSiento* corpus, a new Spanish resource made available to the scientific community[2] that can be used in polarity classification tasks for Spanish opinions. Firstly, we explain the procedure followed for generating this corpus. Then, main statistics of the corpus are shown.

A similar approach to Kamvar and Harris [20] has been followed to generate the *MeSiento* corpus but only using Spanish tweets. The *WeFeelFine*[3] resource served under the Open Linked Data project[4] was found useful for polarity classification in English [24]. Instead of translating this or other resources, we have generated our one from Spanish micro-blog posts, by retrieving those Spanish tweets that contain the words "*me siento*" (*I feel*). This procedure can be considered a streaming method for knowledge-base

[1]`http://www.google.org/flutrends`

[2]*MeSiento* is freely available by contacting the authors

[3]http://wefeelfine.org

[4]`http://linkeddata.org/`

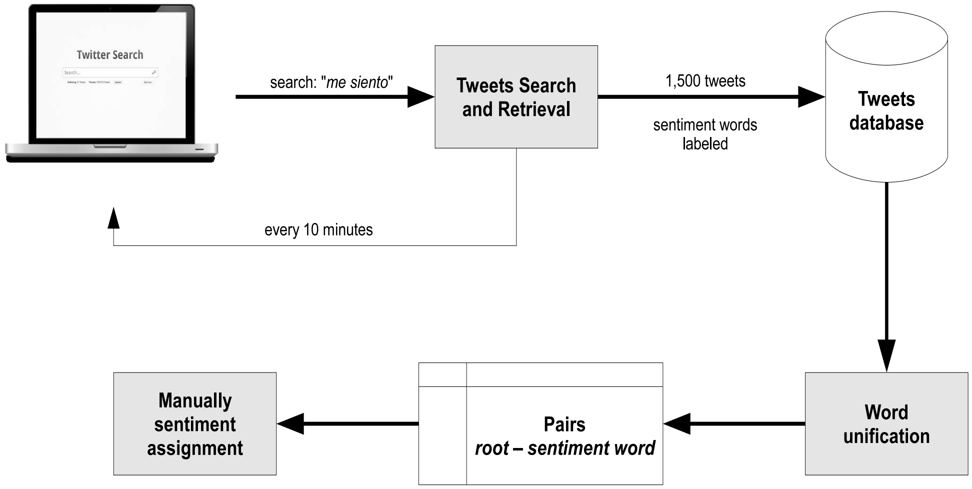

Figure 1: Steps followed to generate the *MeSiento* corpus

generation in polarity classification using social media such as Twitter. The tweets were retrieved during 35 days, between December 2012 and January 2013, collecting a total of 1,863,758 tweets. Figure 1 illustrates the approach carried out to generate the *MeSiento* corpus, which is explained in detail below.

The Twitter API[5] has been used to retrieve all the tweets containing the words "*me siento*" from the continuous stream of posts. This process was launched every 10 minutes and 1,500 tweets were retrieved approximately. Then, those tweets without sentiment information were discarded. This *sentiment information* was established taking into account the words that appear close to the expression "*me siento*", so we tried to detect two types of expressions: "*me siento X*" or "*que X me siento*". The words retrieved within these expressions were labeled depending on whether appearing *before* or *after* the expression. Stop words were also discarded in order to detect comparison sentences such as "*me siento como un X*" (*I feel like X*) or "*me siento tan X*" (*I feel so X*). The stop word list used was composed of these words:

$al|algo|bastante|como|con|de|del|el|en|hasta|$
$la|las|ms|mas|menos|mucho|muy|para|poco|$
$re|tan|un|una|uno.$

If after removing these stop words no word was detected, then we discarded that tweet because we considered that did not contain sentiment information. Finally, the remaining tweets along with the sentiment words detected were stored in a database.

In the second phase we carried out a word unification process. Most of the words detected as sentiment words (or *feelings*) were adjectives, appearing sometimes with different gender (male or female), e.g. "*bueno*" and "*buena*" (*good*). For this reason we performed a manual process to unify similar words but with different gender. In a first step, we selected about 600 most frequent words. Then, the unified form from the male form was manually assigned to the sentiment word, or the same word if the gender of the sentiment word was neutral, e.g. "*antinatural*" (*unnatural*), "*bien*" (*well*) or "*mal*" (*wrong*). Besides, if some sentiment word was misspelled but it was frequent then its unified form was also assigned, e.g. for the word "*maaaaaal*" its unified form was "*mal*". All pairs *unified form - sentiment word* were also stored in the database.

Finally, a sentiment assignment process was manually performed by us (with fully agreement among three annotators), by establishing a positive, negative or neutral sentiment to each unified form. We also discarded those unified forms that could be considered as non sentiment words, such as word in non-Spanish language (*alone*, *crazy*). This is the only work that needs human intervention, though the effort is minimal (the extracted emotions were labeled in less than ten minutes).

The *MeSiento* corpus was analyzed while it was being generated. From the total of 1,863,758 tweets collected, 1,516,184 tweets were not a *retweet* (RT). The number of sentiment words selected was 201, of which 84 were considered as positive and 117 as negative. The total number of different unified forms was 344, while the total number of different words ($positive + negative + neutral$) was 538. Figure 2 shows the number of tweets retrieved per hours.

4. EXPERIMENTAL FRAMEWORK

This section describes the tools employed during the generation of the *MeSiento* corpus, along with the main measure used to evaluate the experiments. Then, we present other corpora used to compare the performance of the proposed corpus.

Regarding the evaluation, we have used *Accuracy* (Acc) as one of the traditional measures employed in text classification (1). Accuracy combines both precision and recall, calculating the proportion of true results (both true positives and true negatives)[28]. In Equation 1, TP (True Positives) are those positive assessments that were correct and

[5] `https://dev.twitter.com/docs/api/1.1`

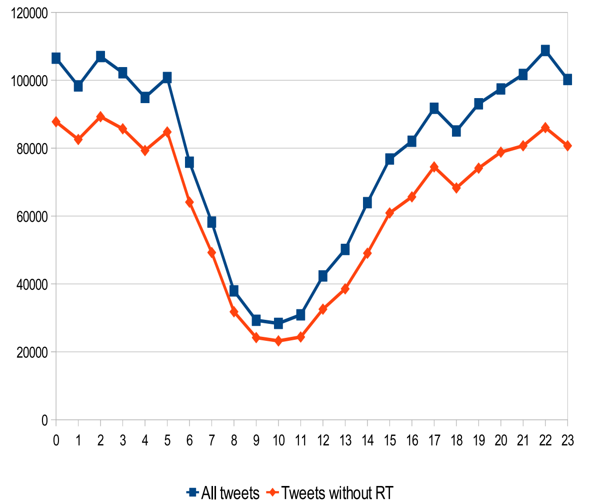

Figure 2: Number of tweets retrieved per hours

TN (True Negatives) are those negative assignments that were also considered negative in by the gold rule.

$$Acc = \frac{TP + TN}{totalassignments} \tag{1}$$

Finally, in order to verify the performance of the proposed corpus regarding the polarity classification task, we decided to compare it with three different Spanish corpora expressing opinions and sentiments: the Spanish SFU Review corpus, the Martínez-Cámara et al. corpus of Spanish tweets and the TASS 2012 corpus.

4.1 The Spanish SFU Review corpus

The SFU Review corpus [30] was firstly generated by collecting reviews from the *Epinions* web site[6]. They applied the Brill's tagger in order to extract only the adjectives found in each review. Then, some adjectives were discarded such as *determiner-like* adjectives (*previous*, *other*) or adjectives that had very low hits after a web search (misspelled adjectives, novel compounds, e.g., *hypersexual*, *club-ready*, *head-knodding*, etc.). The reviews are divided in eight categories (books, cars, computers, cookware, hotels, movies, music and phones), with 25 positive and 25 negative reviews in each category. The classification into positive and negative was based on the "*recommended*" or "*not recommended*" tag that the reviewer provided.

After the SFU Review corpus in English the authors generated its Spanish parallel version, providing a corpus composed of 400 reviews also divided in eight categories: cars, hotels, washing machines, books, cell phones, music, computers, and movies. Each category contains 50 positive and 50 negative reviews, defined as positive or negative based on the number of stars given by the reviewer (1-2=negative; 4-5=positive; 3-star review are not included). These reviews were collected from the *Ciao* web site[7].

[6] `http://www.epinions.com`
[7] `http://www.ciao.es`

4.2 The Martínez-Cámara et al. corpus of Spanish tweets

Martínez-Cámara et al. [23] generated a corpus of tweets in Spanish by using the Twitter Search API[8]. This API allows you to set the language of the retrieved tweets, so the authors used Spanish for collecting the tweets. The authors make use of the emoticons that some users include in their tweets in order to retrieve those tweets expressing opinions or sentiments. They consider as positive those tweets that contain a positive emoticon like "`:)`", and as negative those tweets that contain a negative emoticon like "`:(`". All the tweets were processed by discarding some specific features of Twitter such as *retweets*, mentions to other Twitter users, web links or *hashtags*. Finally, the corpus is composed of 34,634 tweets, of which 17,317 are considered as positive and the other 17,317 are considered as negative.

4.3 The TASS 2012 corpus

The SEPLN (*Sociedad Española para el Procesamiento del Lenguaje Natural*) organization held annually a conference where relevant works related to the Natural Language Processing (NLP) field are presented. In the SEPLN 2012 conference, a new workshop on sentiment analysis called TASS (*Taller de Análisis de Sentimientos de la SEPLN*) was held. The organizers of this workshop provided a corpus of 70,000 tweets written in Spanish, by nearly 200 well-known personalities and celebrities of the world of politics, economy, communication, mass media and culture, collected between November 2011 and March 2012. These tweets were retrieved from users with a diverse nationality, including people from Spain, Mexico, Colombia, Puerto Rico, USA and many other countries, reaching therefore a global coverage in the Spanish-speaking world. Detailed information about this corpus is available on the TASS 2012 web site [9].

5. POLARITY CLASSIFICATION APPROACH

The main idea behind it is to represent each tweet to be classified as a ranked vector of feelings. Then, a final polarity value is calculated from this vector. To this end, we have generated a collection of 200 documents, corresponding to the most frequent feelings collected. Thus, for each feeling, there exists a document containing thousands of tweets. These documents are indexed for further retrieval using the Lucene[10] engine (version 3.6.1 with its default configuration), as shown in Figure 3.

Once the index is generated, we take each tweet as a query and send it to the search engine, retrieving the closest feelings. Finally, from the ranked list of feelings, the final polarity of the tweet is computed based on the polarity value manually assigned to each feeling. The only parameter that has to be specified is the number of results to be used before averaging, which determines the number of feelings to be taken into account when computing the polarity according to one of the two possible equations defined below. For our experiments, this value has been set to 100. The polarity classification process is graphically shown in Figure 4.

[8] `http://dev.twitter.com/doc/get/search`
[9] `http://www.daedalus.es/TASS/corpus.php`
[10] `http://lucene.apache.org`

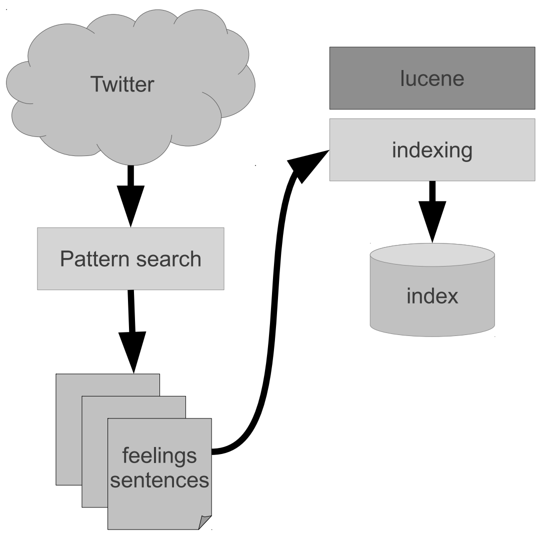

Figure 3: Indexing process

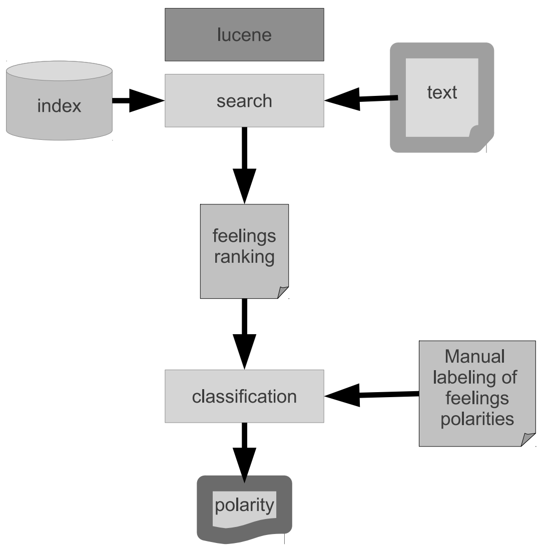

Figure 4: Classification process

The *Ranking Status Value* (RSV) computed by Lucene is also considered when computing the final polarity. By using this ranking value (which reflects a distance between the query tweet and a feeling), we can perform a weighted summatory. Therefore, the polarity proposed formula is defined as follows:

$$p(t) = \frac{1}{|R|} \sum_{r \in R} RSV_r \cdot l_r \qquad (2)$$

where

- $p(t)$ is the polarity of tweet t
- R is the list of retrieved feelings
- l_r is the polarity label of feeling r
- RSV_r is the RSV of the feeling according to Lucene's ranking method

6. EXPERIMENTS AND RESULTS

Several experiments have been accomplished in order to evaluate the *MeSiento* corpus as a knowledge base for Polarity Classification. In our experiments, binary decisions are for the first two data sets, that is, the system only categorizes between positive or negative, without neutral choice. For the TASS corpus, a 3-class (including neutral choice) is taken into account in the final accuracy measurement, to enable comparison with TASS official results. The accuracy values obtained on each of the three studied corpora are shown in Table 1.

Test data	Accuracy	Classes
Spanish SFU Review	**68.25%**	2-class
Martínez-Cámara et al.	64.86%	2-class
TASS 2012	41,04%	3-class

Table 1: Accuracy values obtained for each corpora studied by applying the proposed method

Test data	Accuracy	Classes
Spanish SFU Review	65.00%	2-class
Martínez-Cámara et al.	**74.27%**	2-class
TASS 2012	**50,24%**	3-class

Table 2: Comparison with other works

Table 2 shows the results officially obtained by the authors or participants for each of the three studied corpora, in order to establish a feasible comparison. Should be noted that for the official TASS result we have calculated the average value of accuracy between all the participants. As can be seen in Table 2, the results obtained using the proposed method are promising. Regarding the SFU Review corpus, our approach improves 5% the average result obtained by Taboada and Grieve. However, as regards the other two corpora, no improvement was achieved but the low difference obtained encourages us to continue working in this direction.

7. CONCLUSIONS AND FURTHER WORK

The system proposed can be considered a first attempt in Spanish polarity classification with little effort in the way that tweets are extracted and processed. The obtained results look promising and motivate us to continue the exploration of this approach, but many issues remain opened, like lexical normalization, so informal expressions could be better conflated [15], jargon properly represented, and emoticons also considered. Besides, testing on other languages, filtering of the indexed tweets by means of term-to-class association measures (e.g. Log-Likelihood Ratio), or negation treatment [35], among other questions.

Although the results obtained are not the best compared to those obtained on the same corpora tested, it should be noted that this is an unsupervised approach and, despite this, the accuracy levels achieved are close to those levels obtained with well-known supervised algorithms. Besides, the language independence shown by the method can be interesting in those domains where resources in certain languages are not always available. Therefore, we plan to make available this corpus and the ones to come in other languages under the Open Linked Data project, by means of a REpresentational State Transfer (REST) API.

Another main drawback (though small in terms of effort) is the manual annotation of most frequent emotion extracted in order to create the index. Anyhow, additional resources could be used here, like translated versions of SentiWordNet [2] or other concept-based indexes, like SenticNet [6]. The latter approach (built on top of other linguistic resources) also suggests a way to both data cleansing and conceptual expansion of the collected information that may lead to most effective indexing. The representation of emotions beyond the classical positive-negative dimension is also under study, by modeling the extracted emotions into the four dimensions proposed by the Hourglass of Emotions [7].

We plan to refine the extraction method of feeling-related tweets, always with this independence as main goal, and to test it with other corpora. Current approaches on Multilingual Sentiment Analysis [4] rely on the translation of lexicons or resources. In our case, a crawler of emotional publications by means of simple regular expression matching would allow us to target any other languages.

Finally, to perform a better comparison of our system with current solutions, supervised and unsupervised state-of-the-art methods should be applied on the same tested data sets.

8. ACKNOWLEDGEMENTS

This study has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), TEXT-COOL 2.0 project (TIN2009-13391-C04-02), ATTOS project (TIN2012-38536-C03-0) from the Spanish Government. Also, this study is partially funded by the European Commission under the Seventh (FP7 - 2007-2013) Framework Programme for Research and Technological Development through the FIRST project (FP7-287607). This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

9. REFERENCES

[1] S. Asur and B. A. Huberman. Predicting the future with social media. *CoRR*, abs/1003.5699, 2010.

[2] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In N. C. C. Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may 2010. European Language Resources Association (ELRA).

[3] A. Balahur, J. M. Hermida, and A. Montoyo. Building and exploiting emotinet, a knowledge base for emotion detection based on the appraisal theory model. *IEEE Transactions on Affective Computing*, 3(1):88–101, 2012.

[4] A. Balahur and M. Turchi. Multilingual sentiment analysis using machine translation? *WASSA 2012*, page 52, 2012.

[5] C. Banea, R. Mihalcea, and J. Wiebe. A bootstrapping method for building subjectivity lexicons for languages with scarce resources. In *LREC'08*, pages –1–1, 2008.

[6] E. Cambria, C. Havasi, and A. Hussain. Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis. In *FLAIRS Conference*, 2012.

[7] E. Cambria, A. Livingstone, and A. Hussain. The hourglass of emotions. In A. Esposito, A. Esposito, A. Vinciarelli, R. Hoffmann, and V. Müller, editors, *Cognitive Behavioural Systems*, volume 7403 of *Lecture Notes in Computer Science*, pages 144–157. Springer Berlin Heidelberg, 2012.

[8] H. A. Carneiro and E. Mylonakis. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical infectious diseases*, 49(10):1557–1564, 2009.

[9] H.-H. Chen, C.-C. Lin, and W.-C. Lin. Building a chinese-english wordnet for translingual applications. 1(2):103–122, June 2002.

[10] M. Cheong, S. Ray, and D. Green. Large-scale socio-demographic pattern discovery on microblog metadata. In *Intelligent Systems Design and Applications (ISDA), 2012 12th International Conference on*, pages 909 –914, nov. 2012.

[11] G. De Francisci Morales, A. Gionis, and C. Lucchese. From chatter to headlines: harnessing the real-time web for personalized news recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM '12, pages 153–162, New York, NY, USA, 2012. ACM.

[12] A. F. Dugas, Y.-H. Hsieh, S. R. Levin, J. M. Pines, D. P. Mareiniss, A. Mohareb, C. A. Gaydos, T. M. Perl, and R. E. Rothman. Google flu trends: Correlation with emergency department influenza rates and crowding metrics. *Clinical Infectious Diseases*, 54(4):463–469, 2012.

[13] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA, 1998.

[14] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, Nov. 2008.

[15] B. Han and T. Baldwin. Lexical normalisation of short text messages: Makn sens a# twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pages 368–378, 2011.

[16] B. Han, P. Cook, and T. Baldwin. Lexical normalization for social media text. *ACM Trans. Intell. Syst. Technol.*, 4(1):5:1–5:27, Feb. 2013.

[17] M. Hao, C. Rohrdantz, H. Janetzko, U. Dayal, D. Keim, L. Haug, and M.-C. Hsu. Visual sentiment analysis on twitter data streams. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 277–278, 2011.

[18] G. G. Heredero, S. Bandyopadhyay, and A. Laha. Extraction of contextual information from medical case research report using wordnet. In *Proceedings of the Fourth Annual ACM Bangalore Conference*, COMPUTE '11, pages 13:1–13:8, New York, NY, USA, 2011. ACM.

[19] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and*

social network analysis, WebKDD/SNA-KDD '07, pages 56–65, New York, NY, USA, 2007. ACM.

[20] S. D. Kamvar and J. Harris. We feel fine and searching the emotional web. In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, pages 117–126, New York, NY, USA, 2011. ACM.

[21] A. Lohk, K. Vare, and L. Võhandu. First steps in checking and comparing princeton wordnet and estonian wordnet. In *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*, EACL 2012, pages 25–29, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

[22] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Ureña López, and A. Montejo-Ráez. Sentiment analysis in twitter. *Natural Language Engineering*, FirstView:1–28, 1 2013.

[23] E. Martínez-Cámara, M. A. García-Cumbreras, M. T. Martín-Valdivia, and L. A. Ureña-López. Detecting polarity in Spanish tweets. *Procesamiento del Lenguaje Natural*, 47, 2011.

[24] A. Montejo-Ráez. Wefeelfine as resource for unsupervised polarity classification. *Procesamiento del Lenguaje Natural*, 50, 2013.

[25] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.

[26] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pages 79–86. Association for Computational Linguistics, 2002.

[27] D. Rao and D. Ravichandran. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 675–682. Association for Computational Linguistics, 2009.

[28] F. Sebastiani. Machine learning in automated text categorization. *ACM Comput. Surv.*, 34:1–47, 2002.

[29] I. Stavrakantonakis, A.-E. Gagiu, H. Kasper, I. Toma, and A. Thalhammer. An approach for evaluation of social media monitoring tools. *Common Value Management*, pages 52–64, 2012.

[30] M. Taboada and J. Grieve. Analyzing appraisal automatically. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*, 2004.

[31] P. D. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *ACL*, pages 417–424. ACL, 2002.

[32] R. Valitutti. Wordnet-affect: an affective extension of wordnet. In *In Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 1083–1086, 2004.

[33] P. Vossen. *EuroWordNet: a multilingual database with lexical semantic networks*. Kluwer Academic, 1998.

[34] X. Wang, M. S. Gerber, and D. E. Brown. Automatic crime prediction using events extracted from twitter posts. In *Proceedings of the 5th international conference on Social Computing, Behavioral-Cultural Modeling and Prediction*, SBP'12, pages 231–238, Berlin, Heidelberg, 2012. Springer-Verlag.

[35] M. Wiegand, A. Balahur, B. Roth, D. Klakow, and A. Montoyo. A survey on the role of negation in sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 60–68. Association for Computational Linguistics, 2010.

[36] L. Zhang, Y. Jia, B. Zhou, and Y. Han. Microblogging sentiment analysis using emotional vector. In *Cloud and Green Computing (CGC), 2012 Second International Conference on*, pages 430 –433, nov. 2012.