location where products are purchased; Price, for texts that refer to the cost, value or price of the product; Promotion, for text with references to special offers and campaigns; Quality, for texts that refer to the quality, performance, and characteristics that affect user experience; Sponsorship, for texts that refer to awards, competitions, and events that are organized, endorsed or supported by the brand; Support, for texts referring to customer support services and Warranty, for texts with references to postpurchase services.

The following three examples and their intended labeling give a hint about the complexity of the task.

(1) *Me compré un BRAND*[2] *I30 hace 3 años y todo es perfecto, no lo cambio por nada del mundo.* (I bought a BRAND I30 3 years ago and everything is perfect, I would not change it for anything.) QUALITY.

(2) *No hay autos que me parezcan más feeeeeos que el BRAND A147 y el BRAND 3cv ??* (There are no cars that look to me uglier than BRAND 147 and BRAND 3cv ??) DESIGN.

(3) *Cuando lo compre lo pedi con la alogena de agencia y fue un fraude solo me sirvio por 3 meses y no prende y me cobraron 187 por la alogena y afuera sale mas barata* (When I bought it I asked it with halogen agency and it was a fraud, it only worked for 3 months and it does not turn on and they charged me 187 for halogen and out it is cheaper) PRICE & QUALITY.

In section 2 we present a review of related research, in section 3 we describe the classification system; in section 4 the evaluation experiments; in section 5 the results of the evaluation are reported; in section 6, we discuss the results, and finally conclusions are presented in section 7.

2 Related work

Vázquez et al. (2014) presented an experiment for classifying similar user-generated texts. They used Decision Trees as method for building the classifiers and a Chi-squared selection method for building the BoW. The MM categories addressed and the classifiers results in a 10 fold cross-validation testing experiment are shown in Table 1.[3]

	P	R	F1
Point of sale	0.55	0.41	0.47
Price	0.67	0.35	0.45
Custom.Service	0.38	0.04	0.06
Advertisement	0.88	0.8	0.84
Quality	0.56	0.18	0.27
Design	0.67	0.3	0.41
Promo	0.62	0.32	0.42
Sponsor	0.83	0.37	0.51

Table 1: Classes and results of Vázquez et al. 2014, for a similar Spanish dataset

A similar task to MM classification is aspect identification, one of the subtasks of Aspect Based Sentiment Analysis (ABSA) that was evaluated in the framework of SEMEVAL (Pontiki et al., 2014). Used texts were laptops and restaurant reviews. The goal was to identify product aspects mentioned in the review, for instance if a customer was talking about the quality, price and service of a restaurant.

Most teams that participated at SemEval-2014 ABSA used SVM based algorithms. The NRC-Canada system (Kiritchenko et al., 2014), which achieved the best scores (88.57 % F1 and 82.92 % Acc), used SVMs with features based on various types of n-grams and lexical information learned from YELP data. Other systems equipped their SVMs with features that were a linear combination of BoW and WordNet seeds (Castellucci et al., 2014)[4], or they used aspect terms extracted using a domain lexicon derived from WordNet and a set of classification features created with the help of deep linguistic processing techniques (Pekar et al., 2014), or they only used BoW features (Nandan et al., 2014). Similarly, Brun et al. (2014) used BoW features and information provided by a syntactic parser to train a logistic regression model that assigned to each sentence the probabilities of belonging to each category. Other teams used the MaxEnt model to build classifiers, where only a BoW was used as features (Zhang et al., 2014) or used BoW and *Tf-idf* selected features (Brychcín et al., 2014). Liu and Meng (2014) developed a category classifier with the MaxEnt model with the occurrence counts of unigrams and bigrams

[2] All brands are anonymized in this paper.

[3] However, comparison with their results is not possible because of the different corpus.

[4] In the unconstrained case, they used an ensemble of a two binary SVM-based classifiers and achieved 85.26% F1 and 76.29% accuracy.

words of each sentence as features. Other participating teams only employed WordNet similarities to group the aspect terms into categories by comparing the detected aspect terms either against a term (or a group of terms) representative of the target categories (García Pablos et al., 2014) or against all categories themselves (Bornebusch et al., 2014). Veselovská and Tamchyna (2014) simply looked up the aspects' hyperonyms in WordNet. This approach, however, had many limitations and the systems that used it were ranked in the last positions. And finally, the SNAP system (Schulze et al., 2014) proposed a hybrid approach that combined a component based on similarities between WordNet synsets of aspect terms and categories and a machine learning component, essentially a BoW model that employed multinomial Naive Bayes classifier in a one-vs-all setup.

3 System description

Our system was based on a basic text classification approach. In this method, sentences are represented as Bag of Words (BoW) and a classifier is trained on these representations to recognize every particular class. We built a classifier for each of the nine categories listed in the previous section, because, as we have already shown in (3), texts may belong to more than one category and a multiclassifier would assign only one label.

Therefore, every text is send to nine classifiers to get one or more tags. Many of the texts in the corpus (up to 74% of the whole corpus) are not consumer's statements (herein after NCs) but news or advertisements. These should not be considered business indicators and therefore the nine built classifiers must identify these NC texts by not assigning them any label (see some examples in 4 and 5).

The contributions of our system design include the following developments upon this basic approach. First, we used a reduced BoW for handling vector sparsity, because a BoW with all the vocabulary for such short texts would deliver a very sparse vector with most of the features having 0 as value. Therefore, a selection of 1000 words from the training corpus was made for representing sentences. However, such a reduced BoW could limit the coverage of the system, since it is likely that these selected words do not occur in every text to be classified. In order to enlarge the coverage, a list of synonyms and related words was added to every word of the selected BoW so that when converting the sentence into the feature vector, the occurrence of the selected word or its synonyms were considered a positive feature. In this way, the reduced dimensionality of the vector is maintained, while the number of words that were taken as features was enlarged. Note that we used binary vectors, because frequency effects were not expected to occur in such short texts. Second, we experimented with using Word Embeddings (WEs) and vector space-based measures (Mikolov et al., 2013) to automatically produce the lists of synonyms and related words. In what follows, we explain these contributions in detail.

3.1 Feature selection

The BoW representation of texts has been successfully used for document classification (Joachims, 2001). However for short text classification, this approach delivers very sparse vectors, which are not useful for classification purposes. Different techniques have been devised for vector dimensionality reduction, among these, the ones based on statistical feature selection according to an observed training dataset. In our experiment, we used Adjusted Mutual Information, AMI (Vinh et al., 2009), and chi-squared test to select the words for representing sentences. While AMI, and in general Mutual Information based measures, are known to be useful to identify relevant features, they are biased towards infrequent words. To compensate this bias, we combined it with chi-squared selected ones. Thus, our system first ranks the best candidates in two separated lists, each using a different measure. Then, the two lists are joined into a new one by summing the AMI and chi-squared scores[5]: if a word is ranked 3rd by AMI and 5th by chi-squared, in the joined list it will be the 8th. A single BoW is used for all the classifiers.

3.2 Coverage of word lists

As explained before, an initial BoW was enriched with synonyms and related words, since our intuition was that it is unlikely that every text to be classified contains only some of these words. For instance, texts might contain the word 'costly', present in the BoW, but it

[5] In case of tie, results are ordered alphabetically.

might also contain 'expensive' instead, or even related words like 'cheap' or 'bargain', also useful for the purpose of classifying the sentence in the Price MM category.

Many systems facing this recall problem (see section 6 on related work) rely on external resources like WordNet to supplement initial lists with synonyms by implementing a lexical lookup or a database query component. While technically, this is an efficient and easy solution, its main drawback is that language resources such as WordNet are still missing for many languages. Moreover, these resources do not normally contain the lexica that occur in social media, including abbreviations, slang, etc. (Taboada et al., 2011).

We propose using distributional vector space models and WEs to find relevant synonyms and related words. WEs have demonstrated to perform well to find semantically related words. There are several methods to measure word similarity, but cosine distance has become one of the standard measures (Levy and Goldberg, 2014). Given two vectors as obtained with `word2vec` (Mikolov et al., 2013), related words are obtained by maximizing the function (1), where *cos*θ can be assessed with (2).

$$\arg\max_{a,b \in V} \left(\cos(a, b^{(n)})\right) \qquad (1)$$

$$\cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \cdot \sqrt{\sum_{i=1}^{n} y_i^2}} \qquad (2)$$

Where a and b are WE vectors, V is the vocabulary of the vector space, and n are the nearest candidates n = 1, 2, 3, ... Examples of related words are shown in Table 2 for English and in Table 3 for Spanish. Related words are added according to their cosine distance. A parameter allows selecting the number of closest n words to be added to each feature.

store	grocery shop retailer supermarket retail
bad	terrible poor horrible awful good nasty unfortunate atrocious faithed
wow	hey betcha yeah ah-ha whoa kidding awesome
taste	sweetish fruitiness tartness piquancy flavour flavourful sourness semi-sweet

Table 2: Resulting similar words for EN

teléfono	telefónico móvil telefonía push-to-talk vídeo_llamada pda's smartphone banda_ancha_móvil
respuesta	responder contestación pregunta contestar estímulo explicación anti_viral provocación formular
chocolate	galleta caramelo helado mantequilla golosina praliné bizcocho merengadas anisete

Table 3: Resulting similar words for ES

4 *Methodology*

In this section we describe the experiments carried out to evaluate our system. These experiments focused on two major issues:

(i) The unbalanced distribution of the dataset.

(ii) The validation of the hypothesis that using semantically related words was to increase in particular the classifier coverage.

For the experiments, we trained a Sequential Minimal Optimization for Support Vector Machines (SMO, as implemented by Weka, Hall et al., 2009). The BoW was produced as explained in section 2.1 using the training dataset as shown in Table 5.

As for the corpus, we used the corpus provided by a marketing company with 24,500 manually annotated texts for Spanish and 8,400 for English. The texts were basically tweets, but also microblogs and other social media materials were included. Selection of texts was based on mentions to particular brands. Selected brands represented five different business sectors: automotive, banking, beverages, sports and retail. In Table 4 we can see the distribution of the Spanish (ES) and English (EN) datasets used for the experiments.

Class	ES #	EN #	Class	ES #	EN #
NC	26073	6675	promotion	639	364
advertisement	2131	680	quality	1407	679
design	1237	250	sponsor	348	133
point of sale	925	263	support	680	786
price	1366	422	warranty	86	18

Table 4: ES and EN Datasets distribution

Texts were processed as follows. First, they were cleaned eliminating urls, hashtags, and rare characters. Second, texts were tokenized and lemmatized using Freeling 4.0 (Padró and Stanilovsky, 2012). Stop words were eliminated before assessing the combined AMI+Chi-Squared rank explained in section 2.1. Note that brand names were also ignored and were never selected for the BoW. Once obtained the list of selected words, another module read texts and converted them into 1000 dimension vectors.

4.1 Vector Space Model

As explained, `word2vec` was used to create the vector space model to extract WEs. A 10 window `word2vec` Skip-Gram with negative sampling was trained with the following corpora: we used a Wikipedia dump[6] and the social media datasets totaling 495M words for Spanish and 636M words for English. Both corpora were cleaned and lemmatized as explained before. Other parameters were: algorithm SGNS, 300 dimensions, context window = 10, subsampling $t=10^{-4}$, context distribution smoothing = 0.75, and 15 iterations.

4.2 Training the classifiers

As Table 4 shows, the dataset distribution has an important number of NCs texts. In order to determine the best distribution of positive and negative training examples for such an unbalanced dataset, a preliminary experiment was carried out. The basic issue was to tune the classifiers in order to prevent that they only recognize NCs, which were the majority. The experiment showed significant averaged improvement from 0.756 for the 1 positive-1 negative dataset to 0.811 for the 1 positive-3 negatives dataset.

Therefore, the following experiments followed this distribution where negative samples were randomly selected among the other classes –taking care of the possibility of a particular sample belonging to two or more classes—and NCs. A 70% of the corpus described before was used for training. The remaining 30% was used for testing. Table 5 shows the final number of samples used for training.

The following experiments were carried out to compare a BoW baseline to our proposal. The baseline was made with just words selected as features by the AMI-Chi-squared filter. Five experiments were carried out with different number of related words. In the next section we present only the best results obtained by adding nine related words.

	ES		EN	
	positive	negative	positive	negative
ad	1546	4360	680	2037
design	787	2482	341	807
point of sale	843	2609	263	861
price	874	2606	422	1264
promo	460	1527	364	1144
quality	1052	3125	679	1910
sponsor	187	677	133	438
support	498	1593	786	2191
warranty	81	308	18	70

Table 5: Training test set distribution

5 *Results*

The following results were obtained in two scenarios: a 10 fold cross-validation (Tables 7 and 9) and with the held-out test set (Tables 8 and 10) that was a 30% of the dataset described in section 3. Accuracy is quoted to assess the overall performance of the classifiers.

Significant differences are indicated in bold.

In Table 6 the actual distribution of the held-out test set is described. Note that the held out test sets maintain the distribution of the original datasets with many more negative cases than positive ones.

	ES		EN	
	positive	negative	positive	negative
ad	585	9699	238	3955
design	450	9834	73	4120
point of sale	82	10202	136	4057
Price	492	9792	104	4089
promo	179	10105	96	4097
quality	355	9929	240	3953
sponsor	161	10123	26	4167
support	182	10102	412	3781
warranty	5	10279	5	4188

Table 6: Held out dataset distribution

	ES BASE 10F			ES 9 10F		
	P	R	Acc %	P	R	Acc %
ad	0.829	0.592	86.1	0.79	**0.643**	86.1
design	0.816	0.582	86.7	0.72	**0.620**	85.0
p. of sale	0.711	0.604	84.3	0.698	0.633	84.3
price	0.823	0.576	86.2	0.748	0.597	84.8
promo	0.79	0.483	85.0	0.661	**0.546**	82.9
quality	0.674	0.501	81.3	0.634	0.512	80.2
sponsor	0.789	0.481	85.9	0.724	**0.604**	86.4
support	0.745	0.641	86.2	0.7	0.657	85.1
warranty	0.833	0.679	90.4	0.797	0.679	89.7

Table 7: Detailed results for 10 fold cross validation evaluation of baseline vs. 9-added related words for every class, Spanish dataset

[6] Snapshots of 19-03-2016.

	ES BASE HO			ES 9 HO		
	P	R	Acc %	P	R	Acc %
ad	0.456	0.676	93.5	0.350	0.724	90
design	0.269	0.471	92	0.215	**0.562**	89.1
p. of sale	0.028	0.39	88.8	0.022	**0.451**	84.1
price	0.372	0.43	93.8	0.243	**0.495**	90.2
promo	0.182	0.474	95.3	0.083	0.508	89.4
quality	0.164	0.411	90.7	0.099	0.408	85.1
sponsor	0.086	0.267	94.4	0.072	**0.434**	90.4
support	0.145	0.516	93.7	0.120	0.554	92.0
warranty	0.072	**0.8**	99.4	0.012	0.6	97.7

Table 8: Detailed results for held-out test set validation evaluation of baseline vs. 9-added related words for every class, Spanish dataset

	EN BASE 10F			EN 9 10F		
	P	R	Acc %	P	R	Acc %
ad	0.92	**0.**828	93.8	0.898	0.813	93
design	0.703	0.484	82.9	0.665	**0.564**	82.9
p. of sale	0.573	0.802	81.4	0.624	0.825	84.2
price	0.805	0.697	88.1	0.758	0.699	86.8
promo	0.912	0.797	93.2	0.849	0.805	91.8
quality	0.666	**0.601**	81.6	0.66	0.58	81.1
sponsor	0.683	0.534	83.3	0.613	**0.549**	81.4
support	0.845	0.767	90.1	0.8	0.753	88.5
warranty	0.857	0.333	85.2	0.75	0.5	86.3

Table 9: Detailed results for 10 fold cross validation evaluation of baseline vs. 9-added related words for every class, English dataset

	EN BASE HO			EN 9 HO		
	P	R	Acc %	P	R	Acc %
ad	0.5	0.84	94.4	0.42	0.81	92.7
design	0.05	0.41	87.4	0.06	**0.45**	86.7
p. of sale	0.1	0.84	76.3	0.11	0.82	79.2
price	0.102	0.471	88.4	0.08	0.46	85.8
promo	0.34	**0.802**	95.9	0.24	0.73	94.2
quality	0.21	**0.633**	84.4	0.16	0.56	81.6
sponsor	0.043	**0.73**	89.8	0.03	0.65	87.8
support	0.466	0.762	89.1	0.46	0.74	89.1
warranty	0.03	0.4	98.4	0.006	0.4	92.8

Table 10: Detailed results for held-out test set validation evaluation of baseline vs. 9-added related words for every class, English dataset

6 Discussion

As mentioned earlier, the two main issues were (i) the unbalanced dataset, where the majority of samples do not belong to any class and (ii) how to increase the expected low coverage of the baseline classifiers.

In general, the achieved accuracy shows that the classifiers could handle the unbalanced dataset quite successfully. They could be tuned as to identify many of the texts containing business indicators despite the majority of NC texts. Nevertheless, the precision decrease in the held out test set experiments showed the difficulties of separating NCs and positive cases. These difficulties are maximized with the high number of NCs to classify. Recall that while in the 10 fold cross-validation experiment, negative examples are three for each positive sample, in the held-out test set the original distribution is maintained. For instance, for the English warranty class, in one experiment there are 2 positive and 7 negative samples, while in the other there are 5 positive and 4188 negative samples. For most of the classes, the error analysis showed that 88% of false positive cases were NCs. See in (4) and (5) two examples of false positives for Advertising and Design classes.

(4) For Sale BRAND: Mustang GT 1969 mustang convertible gt clone see video very solid match 70…

(5) Can't believe my little car has been recalled and taken away! The first weekend I plan to drive down the motorway ??

As for the classifiers coverage, error analysis carried out for the held out test set showed that many keywords were already selected by the combined AMI+Chi-squared method making the extended BoW not contributing to the expected extend and instead adding some noise that lowered precision. In the following examples, we mark in bold words that were already in the reduced BoW and underlined words that were in the extended BoW. In (6) we show a Design false negative case that was finally tagged as NC. In (7) an Advertisement text that got the Quality label. In (8) another Advertisement text that got the Support label.

(6) **Need** these! @BRAND SPINS PLASTIC FROM THE OCEAN INTO <u>AWESOME</u> KICKS. *Design → NC.*

(7) I <u>wonder</u> who's BRAND's agency. Their **billboards** are <u>terrible</u>. *Ad → Quality.*

(8) Saw a **commercial** about @BRAND having faster **service** or **connection** <u>now</u>. N my **phone seemed** to go opposite of what the commercial **said**. <u>Great</u>. *Ad → Support.*

Thus, to add related words and synonyms improved only moderately the coverage of the classifiers. In (9) and (10) we can see some correctly classified examples of Design and Advertisement classes.

(9) I **love** these **crisps**! The "**cheese**" and **onion flavour** is better then **walkers**! *Design*

(10) I understand since I don't pay I have **commercials** in between **songs** but that spokesman for @BRAND is **annoying** as **hell please drop** those **commercials**. *Ad*

Finally, the results showed differences between languages that are related to the fact that the Spanish dataset is larger and results obtained are more reliable than for the English dataset. Thus, English results could be improved with a larger dataset.

7 Conclusions

We have presented a system for classifying short texts into the classes of the Marketing Mix model. The task is approached with a supervised machine learning method which works with a reduced bag of word of 1000 features that are selected via a combined AMI and chi-squared ranking method. Each selected feature is complemented with related words as found by cosine distance measure in a distributional vector space made of word embeddings. The results show the feasibility of the approach that intended to maximize coverage in order to identify as many consumer's statements as possible.

Acknowlegments

This work was supported by the Spanish CIEN project LPS-BIGGER cofunded by the MINECO and CDTI (IDI-20141260) and TUNER project TIN2015-65308-C5-5-R (MINECO/FEDER, UE). We want to thank Alberto Sánchez from Havas Media Group for his support.

References

Borden, N. H. 1964. The concept of the Marketing Mix, *Journal of advertising research* 4:2-7.

Bornebusch, F. G. Cancino, M. Diepenbeck, R. Drechsler, S. Djomkam, A. Nzeungang Fanseu, M. Jalali, M. Michael, J. Mohsen. M. Nitze, C. Plump, M. Soeken, M. Tchambo, and H. Ziegler. 2014. itac: Aspect based sentiment analysis using sentiment trees and dictionaries. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 351–355. Dublin, Ireland: ACL and Dublin City University.

Brun, C., D. N. Popa, and C. Roux. 2014. Xrce: Hybrid classification for aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 838–842. Dublin, Ireland: ACL and Dublin City University, August 2014.

Brychcín, T., M. Konkol, and J. Steinberger. 2014. Uwb: Machine learning approach to aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 817–822. Dublin, Ireland: ACL and Dublin City University, August 2014.

Castellucci, G., S. Filice, D. Croce, and R. Basili. 2014. Unitor: Aspect based sentiment analysis with structured learning. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014*), pages 761–767. Dublin, Ireland: ACL and Dublin City University, August 2014.

García Pablos, A., M. Cuadros, and G. Rigau. 2014. V3: Unsupervised generation of domain aspect terms for aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 833–837. Dublin, Ireland: ACL and Dublin City University, August 2014.

Hall, M., E. Frank, G. Holmes, G., B. Pfahringer, P. Reutemann, and I. H. Witten. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations* 11(1):10-18.

Joachims, T. 2001. A Statistical Learning Model of Text Classification with Support Vector Machines. In *Proceedings of the Conference on Research and Development in Information Retrieval (SIGIR)*, ACM, pages 128-136.

Kiritchenko, S., X. Zhu, C. Cherry, and S. Mohammad. 2014. “Nrc-canada-2014: Detecting aspects and sentiment in customer reviews”. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 437–442.

Levy, O., Y. Goldberg, and I.Ramat-Gan. 2014. Linguistic regularities in sparse and explicit word representations. CoNLL-2014.

Liu, P. and H. Meng. 2014. “Seemgo: Conditional random fields labeling and maximum entropy classification for aspect based sentiment analysis”. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 527–531.

McCarthy, E.J. (1978), Basic Marketing, a Managerial Approach, Sixth Edition, Homewood, Ill.: Richard D. Irwin, Inc.

Mikolov, T., K. Chen, M. Corrado, and J. Dean. 2013. Efficient Estimation of Word Representations in Vector Space. Proceedings of Workshop at ICLR, 2013.

Nandan, N., D. Dahlmeier, A. Vij, and N. Malhotra. 2014. “Sapri:A constrained and supervised approach for aspect-based sentiment analysis,” in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 517–521.

Padró, Ll. and E. Stanilovsky. 2012. FreeLing 3.0: Towards Wider Multi-linguality. In Proceedings of the Language Resources and Evaluation Conference (LREC 2012) ELRA.

Pekar, V., N. Afzal, and B. Bohnet. 2014. “Ubham: Lexical resources and dependency parsing for aspect-based sentiment analysis,” in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 683–687.

Pontiki, M., D. Galanis, I. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos and S. Manandhar. (2014). SemEval 2014 Task 4: Aspect Based Sentiment Analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) Dublin, Ireland: ACL and Dublin City University, August 2014, pages 27-35.

Schulze Wettendorf, C., R. Jegan, A. Körner, J. Zerche, N. Plotnikova, J. Moreth, T. Schertl, V. Obermeyer, V. Streil, T. Willacker, and S. Evert. 2014. “Snap: A multi-stage xml-pipeline for aspect based sentiment analysis”. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 578–584.

Taboada, M., J. Brooke, M. Tofiloski, K. Voll, and M. Stede. 2011. Lexicon-based methods for sentiment analysis. Computational Linguistics, 37, 2, pages 267-307.

Turney, P.D. and P. Pantel. 2010. From frequency to meaning: Vector space models of semantics. Journal of Artificial Intelligence Research, 37, pages 141–188.

Vázquez, S., O. Muñoz-García, I. Campanella, M. Poch, B. Fisas, N. Bel, and G. Andreu. 2014. "A classification of user-generated content into consumer decision journey stages", Neural Networks, 58, pages 68-81.

Veselovská K. and A. Tamchyna. 2014. “U´ fal: Using hand-crafted rules in aspect based sentiment analysis on parsed data”. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 694–698.

Vinh, N. X., J. Epps, and J. Bailey. 2009. Information theoretic measures for clusterings comparison: is a correction for chance necessary? In Proceedings of the 26th International Conference on Machine Learning (ICML'09), pages 1073- 1080. ACM.

Zhang, F., Z. Zhang, and M. Lan. 2014. “Ecnu: A combination method and multiple features for aspect extraction and sentiment polarity classification,” in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, pages 252–258.