Using the length of the speech to measure the opinion

Luigi Lancieri
University of Lille1, France
luigi.lancieri@univ-lille1.fr

Eric Lepretre
University of Lille1, France
eric.lepretre@univ-lille1.fr

***Abstract*—This article describes an automated technique that allows to differentiate texts expressing a positive or a negative opinion. The basic principle is based on the observation that positive texts are statistically shorter than negative ones. From this observation of the psycholinguistic human behavior, we derive a heuristic that is employed to generate connoted lexicons with a low level of prior knowledge. The lexicon is then used to compute the level of opinion of an unknown text. Our primary goal is to reduce the need of the human implication (domain and language) in the generation of the lexicon in order to have a process with the highest possible autonomy. The resulting adaptability would represent an advantage with free or approximate expression commonly found in social networks environment.**

Keywords—Opinion-mining; unsupervised learning;

I. INTRODUCTION

In the last decade there has been an increasing effort in the linguistic and data-mining community to address the question of the computation of the opinion from a textual content. Opinion mining is often view as a sub-field of sentiment analysis that, as the discourse analysis or the linguistic psychology, seek to evaluate affective state through the analyze of natural language. Nevertheless, many researchers define sentiment, loosely, as a negative or a positive opinion [20, 22]. This relatively new field is not only promising from a scientific point of view but also because of its large possible applications. The challenges are linked to the huge quantity of data available. Thus, applications such as business intelligence, trend forecasting or recommendation systems would take benefits from opinion mining.

The basic principle generally starts with the necessity to use or to generate a lexicon that makes a link between words and their opinion value. Then, this lexicon will be used to rate a text by combining the opinion value of each of its words. Unfortunately, the generation of this lexicon is not obvious because of the complexity of the language rules or of the diversity of the modes of expression. Indeed, the language is alive, evolves and is subject to all kinds of exceptions or common mistakes. It is not rare to find sentences with mixed opinions, with several sources (e.g. quotation of other persons) or several targets of the opinion (e.g comparison of products or features). The literature is full of examples of expressions where the same word can be interpreted differently depending on its use in the sentence. All theses cases introduce biases in the opinion interpretation and reduce the performance of a computational evaluation. Denecke shows, for example; that Senti-WordNet (lexicon with opinion labels) has difficulties to evaluate news articles that are generally wrote in central style. In such a case, the accuracy of the classification stood to 40 % [9]. It is even more difficult in social networks environment that are heterogeneous and noisy by nature. The freedom of expression that tends to lead to abbreviations or malformed sentences does not fit to standard lexicons. It is important to point out that not only the generation of the lexicon is tough from a linguistic point of view but also because a lexicon is highly context dependent. In other words, each context in a specific language need a dedicated effort to solve these not obvious problems. In the same way existing lexicons cannot be easily transposed in other languages or other contexts.

In consequence, facilitate the production of the lexicon seems to be a major research issue. Apart manually generated lexicons, there are several, more or less, automated solutions. Often using machine learning techniques, researchers attempt to make easy the lexicon generation and try to offer a better adaptivity to these multiple variations. But, how far can we go into this simplification?

In this paper, we present our contribution that takes profit of the natural asymmetry of the expression of opinions. In short, this psycholinguistic feature of the human behavior makes that negative narrations are longer than positive ones. This is observable in several situations of free expression, as on web sites dedicated to e-commerce, when the users give their feeling on their buy or on the quality of the merchant. Probably because there is less to say when all is going well than when it is going bad, negative posts are statistically longer than positive ones. We find that, with enough of such independent narrations, the length of the text can be used to automatically differentiate positive from negative vocabulary and generate a list of words with polarity tags. Such "lexicons" can then be used to evaluate the opinion level of an unknown text. Since the need of prior knowledge is limited, the human involvement in the generation process is very low and do not need to be technically or linguistically specialized. This allows to create a lexicon as frequently as needed, for a specific domain or language. In order to validate our theory, we collected consumers textual feedback and their associated stars rating. First, we generate the lexicon with a first subset of the users' comment alone. The stars rating are not used in the lexicon generation phase. Then we compute the opinion value of the second subset and we compare the result with the user stars rating. The rest of this paper is organized in 4 other sections where we first develop a state of the art of opinion mining. Then we present our proposal and the associated results and finally we discuss its perspectives and limitations.

II. State of the art

Opinion mining techniques can be roughly segmented in two categories as they are bottom up or top down. Even if a lexicon is always needed, the way to create it can widely differs. The first category needs the most prior knowledge and start from existing lexicons often manually adapted to include sentiment knowledge tags. The second approach uses a series of textual documents that have been globally annotated. An example is a 5 lines long text of a customer comment provided with a 4 stars rating. These couple of comment / rating are used to infer the polarity of words and to generate a reference lexicon containing words and their polarity. These two opposed approaches have also been combined.

A. Lexicons generation

In this study, we define a lexicon as a list of words with one or several language attributes. This can include a basic list of words with polarity tags or a more complex dictionary with grammatical and sentiment class attributes. The polarity or the affect annotations can be added in several manners but in all cases it needs prior knowledges.

Most of the time, sentiment based lexicons have been manually constructed by extending general purpose lexicons associating words with affects categories and degree of relatedness with the categories [26, 8]. The probably well known example if that of the public domain Princeton WordNet lexicon [21] that have leads to several other versions. The original WordNet is now available in around 80 languages but it is rather static and difficultly open to new languages, to emerging words or to multiple domains. As examples of sensitive lexicons extended from WordNet, we can mention WordNet-Affect that contains labels linking words with several affective categories [25] or Senti-WordNet that adds polarity and subjectivity labels [13]. WordNet has also been used to automatically generate basic lists of positive and negative terms [16, 15]. Let us also mention, among other examples, the Havard General Inquirer lexicon from both the "Harvard" and "Lasswell" general-purpose dictionaries [24].

The Scientific communities also provide manually annotated databases[1] that can be used as language resources or for studies validation. Other initiatives as MIT media Lab Open Mind Common Sense focus on the build of a large knowledge from the contributions of thousands of people across the Web. Basic facts including emotional ones are provided by users (e.g "The sun is very hot"). Such sentences are analyzed in order to extract concepts, ontologies or basic positive-negative lists of words (see also Cyc.com), [19, 27]. In other cases, resources manually rated such as movies, products or merchant rating available on customer opinion web sites are also often used (CNET, Ebay, TripAdvisor, IMDB). The idea is here to use a machine learning algorithm in order to extract a link between words and the rated comment and predict the opinion of an unrated text [7, 14].

B. Identifying the polarity of words

The automation of the lexicon generation involves the use of heuristics that, in short, provide to the algorithm a part of the human expertise. Thus, the identification of the polarity of words can be done using more or less complex methods. In bag of words, terms are considered as independent parts of speech. Elementary grammatical features of words that are known to have polarity values (adjectives, adverb, negation, intensifier, diminisher) are used to separate them. Adjectives or adverbs are then organized in binary classes or associated to a position in a continuum between the two extreme polarities [23]. But, adjectives are not always the best opinion descriptors. For example, in the sentence “There was a lot of noise in this hotel”, all the negative weight resides in the noun “noise”. If we replace it by the nun “facilities”, the sentence become positive. This shows that, beyond adjectives or adverbs, the polarity depends on a larger range of terms individually or in association. Unfortunately, bags of words neglect the position of words in the sentence and only take into account their presence or their frequency. Alternatively, as commonsense and experiments tend to shows it, parts of speech patterns used in the n-gram techniques have a better efficiency. A basic example is the following where a negation (no, not) involves a very different polarity depending on its position in the sentence (this book was not good - no wonder, every one love this book). The co-occurrence of words is also a key criterion. In short, it is assumed that words that have the same polarity are often found together or more or less close in a sentence. This relationship between words can be computed following different techniques as LSA (Latent Semantic Analysis), PMI (Pointwise Mutual information), Chi-Squared test of independence [7].

Recent works exploit the research and the analysis of seeds words in an unknown text. Seeds [27] have an evident polarity (well, good, bad, .) and are used to collect other connoted words. The criterion used to extends these lists can be the level of proximity with the seeds.[12]. A statistic analysis of the collected words helps to refine the lists. This technique needs less expert implication but it requires a good knowledge of the target language and domain. Not only the seeds have to be chosen carefully but in case of domain oriented vocabulary some words may be connoted differently (cold is negative for the evaluation of a restaurant but can be positive in other domains as for describing the quality of a fridge). Actually, the influence of the domain is known as a key issue not only for the opinion mining but also for the knowledge management in general. Consequently, several researches have tackled the sensitivity to the domains or in other words to see how to use a lexicon from one topic to another [4,12].

C. Opinion and length of expression

Even if it is clear that the association of words provides better performances, it is still a question to identify the optimal length of this association [26, 1, 23]. This issue has raised the attention of the community with the spread of micro-blogging platforms, as Twitter, where the size of the message is strongly limited [3, 11,5].

Nevertheless, from our knowledge, the literature does not provides example of study that takes into account the difference of the expression length in order to statistically separate positive from negative vocabulary and generate lexicons. Though, this psycholinguistic feature of the human expression has already been observed by researchers.

Anderson, for example, has stated that unsatisfied customers engage themselves in greater word-of-mouth than

1 TREC (Text Retrieval Conference), NTCIR (NII Text Collection for IR), CLEF (Cross Language Evaluation Forum)

satisfied ones [2]. In the same domain, another study on 5014 reviews of customers shows that positive opinions contain 4 times less words in average than negative ones [17]. In an other context, observing the expression of emotions in computer mediated communication as e-mails, Derks sees that in case of negative situations as conflicts, Emoticons are not enough. This leads to more communication between individuals to solve the problem, whereas in positive situations, a simple smiley can be sufficient [10]. It is also interesting to observe that the performance of automated sentiment miners is better in positive texts than in negative ones (see table 2). This confirms the observation of Gindl [14] and leads to the assumption that positive words are more used in negative opinions that negative words in positive opinions. This would imply that positive opinions are less ambiguous probably due to its conciseness.

III. Methodology

For our experiment, we collected a set of consumers' feedback (comment and stars rating) that we randomly split in a learning (L) and a validation (V) subset. The L subset contains only the comments (i.e. without rating) and is used for the generation of the lexicon. Then we use it to compute the opinion value of the V subset comments (each independently) and we compare the results with the users stars rating. In order to estimate the performance of the process we use the well known precision, recall and f-index ratios for the positive (Pp,Rp,Fp) and the negative class (Pn,Rn,Fn).

At the beginning, the comments of the L set were randomly dispatched into two subsets P0 and N0 that can be viewed as two kinds of bag of words with a loss of organization between words. These two sets will be progressively refined through several iterations where new sets (P1, N1 to Pn,Nn) will be generated from the previous ones. At each iteration *i*, *Pi* and *Ni* will be used to generate 2 steps lexicons *Lpi, Lni.* Let us remark that whereas *Pi,Ni* aggregate the same number of comments and thus can contain several times the same words, the union of *Lpi* and *Lni* sets contains only one occurrence of a word. At the last iteration *Lpn* and *Lnn* are expected to contain words with respectively positive and negative polarities.

At the second step, the frequency of P0 and N0 words are computed in order to generate the Lp0 and Ln0 lexicons. Thus, a specific word will be stored in Lp0, in Ln0 or discarded depending on the difference of frequency it has on the two subsets P0, N0. This step allows to eliminate articles or others neutral words that have a similar frequency in all kind of texts.

```
For each unique word  W_i ∈ (P_i ∪ N_i)
if (Freq(Wi) in Pi) > (2*freq(Wi) in Ni)
        then Wi is stored as unique in Lpi
        else if (Freq(Wi) in Ni) > (2*freq(Wi) in Pi)
                then Wi is stored as unique in Lni
                else Wi is discarded
End for
```

In the third step, the goal is to start the agglomeration of words having the same polarity in the same sets (Lpi or Lni). We take again the L subset and, for each comment, we compute its level of polarity (Pij) with the following formula. For example, if a coment is composed of 13 words where 10 apear in Lp0 and 3 in Ln0. The polarity of this comment would be equal to 0.53 (i.e 10-3/10+3).

$$Pij = \frac{card(Tj\ Words \in Lpi) - card(Tj\ Words \in Lni)}{card(Tj\ Words \in Lpi) + card(Tj\ Words \in Lni)}$$

If Pij is ranging between +1 and K (see below), it is assumed to be of positive polarity. Negative, if it is between -K and -1 and neutral if its between -K and K. Then, if Pij is positive, all words of the text Tj, recognized either in Lp0 or Ln0, are stored in P1. If Pij is negative, the text words recognized are stored in N1. Then the frequency algorithm is processed again but now on P1 and N1 in order to generate Lp1 and Ln1. Statistically speaking, each set (Lp1, ln1) should be a bit more consistent in term of polarity than Lp0 and Ln0, even if this polarity (positive or negative) is not yet known.

In the fourth step, the number of words in P1 and N1 will decide of this polarity. If N1 has more words than P1 then Ln1 is considered as the negative step sub-lexicon and Lp1 the positive one, else Lp1 become the negative one (and is renamed Ln1) and Ln1 the positive one.

All the process from the third stage is iterated until Lpn, Lnn is considered as stable. (experimentally, N=20 iterations was found as enough, see figure 1). Then we built the final consensus lexicons on the basis of the words present in the Z=N/2 (i.e. 10) last step learning lexicons of the same polarity (lp11, lp12,..,lp20 → final Lp; Ln11, Ln12,..,Ln20 → final Ln). Words are kept in the final concensus lexicons if they appear in more than C=Z/2 (i.e. 5) of the Z lexicons.

The value of the K, Z an C parameters are important. The distance [-k, +k] correspond to the neutral polarity proportion. In order to simplify, we considered that each category (positive, negative and neutral) are proportionally equivalent (ie K=0,3). That means that the probability that a text fall in one of these categories is estimated as identical (33%). Actually, this depends of the context and it even seems that, most of the time, negative messages are over represented [17]. K should be different to zero in order to avoid oscillations in the learning process. The N parameter, linked to the need of having a complete learning process, results mainly from the experimental observation. As that can be seen in the figure 1, the iterated process stabilize itself pretty rapidly. Furthermore, it is important to have in mind that since we choose to limit prior knowledge, we have no means, except the use of a heuristic to know when the learning process would be at its optimum. The Z and C parameters as in a vote process define the level of consensus to build the final lexicon.

IV. Results

We applied this process to a set of data in english related to the High Techs domain. Then we compare the reults with other approachs.

A. Experimental validation

In this experiment, we collected 20400 users reviews including comments (68 words average length) and stars rating from the well known *epinion.com* web site. The validation set was randomly composed of 1381 texts (96488 words) from the initial set. The rest of the initial set was used to compose several learning sets in order to evaluate the influence of the size of the learning set. We also evaluate the performance at each iteration of the learning process.

In order to have a synthesis of the performances, we define 3 classes of opinions with their rating (negative: 0 to 2 stars, neutral: 3 positive: 4 to 5 stars). The estimated opinion was computed from the textual comment in order to fit the same scale (ie adapted from the Pij formula). In the ideal situation the user stars rating should correspond to the computed one. The sensitivity of the L size on the performances was evaluated with 8 tested sizes (from 364 to 7053 kBytes). The performances are reported in the two following tables. The first one presents the evaluation ratios with the final consensus lexicons. The second one presents the best values during the 20 steps. We can observe that most of the time the consensus lexicons give the best results (exept in bold in table 1).

Size	Pp	P n	P	Rp	R n	R	Fp	Fn	F
364	65,1	65,2	65,1	62,3	60,7	61,5	63,7	62,9	63,3
1111	***70,3***	***68,9***	***69,6***	***65,9***	***68,7***	***67,3***	***68,0***	***68,8***	***68,4***
2045	64,6	72,2	67,5	76,2	51,4	63,8	69,9	60,0	65,6
2971	67,8	66,6	67,2	63,5	66,8	65,2	65,6	66,7	66,2
3901	***93,2***	***74,9***	***81,4***	***64,3***	***94,8***	***79,5***	***76,1***	***83,7***	***80,4***
5014	86,8	80,7	83,4	74,9	86,7	80,8	80,4	83,6	82,1
6124	74,7	74,8	74,8	73,5	72,3	72,9	74,1	73,5	73,8
7053	85,8	86,0	85,9	84,8	82,3	83,6	85,3	84,1	84,7

Table 1 : Results with the consensus lexicon

Size	Pp	P n	P	Rp	R n	R	Fp	Fn	F
364	65,1	65,2	65,1	62,3	60,7	61,5	63,7	62,9	63,3
1111	77,6	86,4	81,4	83,9	68,7	76,3	80,7	76,5	78,7
2045	64,6	72,2	67,5	76,2	51,4	63,8	69,9	60,0	65,6
2971	65,9	73,3	68,9	75,8	57,1	66,5	70,5	64,2	67,7
3901	89,1	85,4	87,2	81,4	86,6	84,0	85,1	86,0	85,6
5014	85,7	82,6	84,1	79,2	85,0	82,1	82,4	83,8	83,1
6124	74,3	78,6	76,3	77,1	69,1	73,1	75,6	73,6	74,7
7053	84,9	86,7	85,8	85,0	82,5	83,8	84,9	84,6	84,8

Table 2: The best F-index results.

We can see that the size of the learning set is not a clear criterion to have good performances (see in table 2, F-index for 5014 kB, 6124 kB and 7053 kB). It is important to remind that each L set was composed randomly from the original set. This involves that words are not always the same and can cause different lexicons even if the process was run several times with the same set size. This is the main reason that causes important changes in performances even in the final lexicons. This means that the aggregation process that generate these lexicons does not completely catch the optimum performances but succeed to avoid the lower ones. Nevertheless, as shows the following figure (with 3 examples of L size), the learning process converge pretty rapidly with stable results in the last steps. Also, wee see that a size of learning set L from 4 to 7 MB provides reasonable results. In this figure we can also observe the oscillations of the F-index near the inversion of the lexicons polarity (fourth learning step) generally observed at the third iteration. At this point, the lexicons start to become consistent from the polarity point of view.

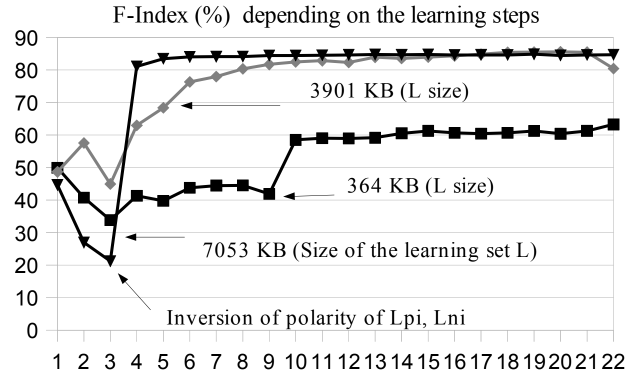

Figure 1: Convergence of the Learning process

B. *Comparison with a Seeds based method*

The use of seeds seems to be the most powerful actual method but the choice of the initial key words (the seeds) may have a strong influence on the final results. In this part of the experiment, we take again the learning set that provide the better results (7053 KB) and the validation set and we build the final lexicon on the basis of the seeds method. In order to show the sensitivity to the initial seeds, we use four examples of 6 seeds (3 positives and 3 negatives) and we compute the precision, recall and F ratio. The first set provides equivalent performance compared with our method. In the second and the third set we changed only one world respectively with negative (set 2) and positive polarity (set 3). In the last set, we change several worlds for each polarity.

- Seeds set 1: super excellent perfectly bad poor expensive
- Seeds set 2: super excellent perfectly bad noisy expensive
- Seeds set 3: super excellent recommend bad poor expensive
- Seeds set 4: good excellent friendly poor noise bad

S. set	Pp	Pn	P	Rp	Rn	R	Fp	Fn	F
1	90	83	86	75	89	82	82	86	84
2	86	73	78	64	85	74	73	79	76
3	83	74	78	63	83	73	72	78	75
4	39	44	42	23	52	38	29	48	40

Table 3: Recall, precision and F-index using the seeds method

The results show clearly that the seeds method is very sensitive to the choice of the worlds even with a careful attention to the context (here, users' comments on hotels). The case of the last set is very interesting. We can observe that even with words that are evidently consistent in polarity and in context, the performances are very bad. The reason is probably due to the representativity level of the seeds words in the learning set. It is important to say that each of these words are present in the learning set but with a different frequency of occurrence. The conclusion could be that the seeds method need either a linguistic and a domain expertise in order to chose the most accurate seeds words or a large size learning set in order to statistically compensate the lack of reprentativity of a specific seed. As landmarks, let us also say that in the 29 studies from 1997 to 2009, reported by Mejova, 19 have more than 80 % of accuracy and 6 more than 90 % [22].

V. DISCUSSION

In this article we show that the psycholinguistic feature of the free individual expression can be used to compute the opinion of texts. This potentially allows a better adaptivity to multiples languages and domains and a better tolerance to approximate expressions or errors (e.g linguistics shortcuts as in twitter). Anyway, our methodology has some limitations. Even if we do not need to have a strong knowledge about the collected texts, we need to know that they contain opinions for a majority of them (customer or blog feedback,..). The other limit is that the inconsistency of the sources, in terms of domains, is difficult to be controlled if we want to limit the human interventions. Thus, a complete blind approach could reduces the performances but this can be enough if the goal is a rough classification. Furthermore, as our first goal was to validate the interest of the speech length heuristic, we spent low efforts on the question of the syntactic analysis which could be improved. Indeed, our basic bag of words strategy can takes benefits from the lot of studies done on this field (n-gram). In terms of perspectives, outside the improvements that we have just evoked, we wish to evaluate the potential of this approach in several practical applications.

REFERENCES

[1] Agrawal, Rajagopalan, Srikant, and Xu (2003). Mining newsgroups using network arising from social behavior. 12th international W.W.W Conference.

[2] Anderson EW,1998, Customer satisfaction and word of mouth - Journal of Service Research, 1998

[3] A. Bermingham, A.F. Smeaton, Classifying sentiment in microblogs: is brevity an advantage? in CIKM '10 Proc. of the 19th ACM international conference on Information and knowledge management, Pages 1833-1836

[4] Blitzer, J., Dredze, M., and Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. Proceedings of the 45th , Annual Meeting of the Association of Computational Linguistics, pages 440–447.

[5] J. Bollen, H. Mao, A. Pepe, Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena, proceedings of the 5th International AAAI Conference Weblogs and Social Media, 2011

[7] J Cogley, Sensing Sentiment in On-Line Recommendation Texts and Ratings - B.A dissertation 2010

[8] Dave, K., Lawrence, S., and Pennock, D. M.. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. Proc. of the W.W.W Conference 2003

[9] K Denecke, Are SentiWordNet scores suited for multi-domain sentiment classification?, in Fourth International Conference on Digital Information Management (2009)

[10] D.Derks A.H. Fischer, A.E. Bos, The role of emotion in computer-mediated communication: A review, Computers in Human Behavior, Vol.24, Issue 3, 2008, Pages 766–785

[11] N.A. Diakopoulos, D.A. Shamma, Characterizing debate performance via aggregated twitter sentiment, in CHI'10 Proc. of the SIGCHI Conference on Human Factors in Computing Systems pp 1195-1198

[12] B.Duthil, F.Trousset, G.Dray, J.Montmain, P.Poncelet, Opinion Extraction Applied to Criteria, Database and Expert Systems Applications, Lecture Notes in Computer Science Vol. 7447, 2012, pp 489-496

[13] Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. Proceedings of the 5th Conference on Language Resources and Evaluation (LREC).

[14] Gindl, S.; Liegl, J. Evaluation of Different Sentiment Detection Methods for Polarity Classification on Web-Based Reviews; 18th European Conf. on Artificial Intelligence (ECAI-2008), Workshop on Computational Aspects of Affectual and Emotional Interaction.

[15] Hu, M. and Liu, B. (2005). Mining and summarizing customer reviews. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing.

[16] Kim, S.-M., Hovy, E. (2004). Determining the sentiment of opinions. Proceedings of the 20th International Conference on Computational Linguistics.

[17] L.Lancieri, E.Lepretre, Sentiment Analysis in Social Web Environments oriented to e-commerce, IADIS Web Based Communities and Social Media conference (WBC 2011)

[19] Liu, H., Lieberman, H., and Selker, T. (2003). A model of textual affect sensing using realworld knowledge. Proceedings of the Seventh International Conference on Intelligent User Interfaces, pages 125–132.

[20] Melville, P., Gryc, W., Lawrence, R. D. Sentiment analysis of blogs by combining lexical knowledge with text classification. Proc. of the Conference on Knowledge Discovery and Data Mining 2009.

[21] G.A.Miller(1995) WordNet, A lexical database for English Communications of the ACM V.38, N.11: 39-41.

[22] Y.Mejova, Sentiment Analysis: An Overview, Technical report, Computer Science Dep. University of Iowa, 2011,

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

[24] PJ Stone, DC Dunphy, MS Smith; The General Inquirer: A Computer Approach to Content Analysis.Oxford, England: M.I.T. Press. (1966). 651 pp..

[25] Strapparava, C. and Vlitutti, A. (2004). Wordnet-affect: and affective extension of wordnet. In Proc. of the 4th International Conference on Language Resources and Evaluation.

[26] Turney, P. D. and Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems (TOIS), 21(4):315–346.

[27] Zhou, L. and Chaovalit, P. (2008). Ontology-supported polarity mining. Journal of the American Society for Information Science and Technology, 69:98–110.