

# Agile sentiment analysis for more responsive public relations

Bojan Cestnik<sup>1</sup>, Alenka Kern<sup>2</sup>

<sup>1</sup> Temida d.o.o., Dunajska cesta 51 &  
Jozef Stefan Institute, Jamova cesta 39  
1000 Ljubljana, Slovenia

<sup>2</sup> The Housing fund of the Republic of Slovenia, public fund  
Poljanska cesta 31, 1000 Ljubljana, Slovenia  
bojan.cestnik@temida.si  
alenka.kern@stanovanjskisklad-rs.si

**Abstract:** The constituting e-government organizations like ministries, agencies, funds and councils often have to deal with severe media attention and sometimes more or less justified criticism. In this paper we present an approach that supports the task of managing public relations (PR) between organizations and the news media. The proposed approach is based on agile sentiment analysis of the questions that organizations receive from the journalists and media. Such analysis is relevant for reviewing past events as well as predicting the future happenings. We demonstrate the utility of the proposed approach by analyzing a set of 298 questions received by a public organization from various media in the period from 2007 till 2012. The results confirm that that by incorporating agile sentiment analysis into regular PR workflow organizations can improve their understanding and control of communication with the media and public.

**Keywords:** e-Government, media analysis, sentiment analysis, knowledge representation.

## 1 Introduction

Representatives of politics and business often address the public to communicate their message or stand on a particular matter. Usually, the decision about the time, place and contents of the message is in the hands of the issuer. However, when the media and general public get interested in the matter, the roles might change. Then, the active role is played by the journalists representing the media, and the representatives of politics or business have to respond to the journalists' inquiries. Many cases of that kind demonstrate how important and delicate such response statements are and how they influence public perception of the matter under consideration.

For example, during the Cyprus crisis in March 2013 the Dutch Finance Minister Dijsselbloem, who as head of the Eurogroup played a key role in the Cyprus negotiations, hastily said that the Cyprus deal will be used as a template for the future solutions of similar Eurozone banking problems [BBC13]. This statement, when restated by the mass media, caused much greater confusion all around the Europe than the crisis itself. Even though he later added a clarification saying that Cyprus was "a specific case

with exceptional challenges", he could not have undone the damage: European markets were already alarmed.

Rational arguments constitute foundations of science, economics and law. Emotions, on the other hand, put flavor to our everyday lives in politics and business. Explanatory models based on reason alone often fail to account for the complexity of reality. An attempt to overcome such limitations by combining rational models and emotional explanatory approach resulted in a new method called sentiment analysis [L10].

Sentiment analysis is a part of social media analysis, which is a mature method for e-Government benchmark [CRI12]. The field of sentiment analysis aims to automatically elicit emotions like happy-sad or positive-neutral-negative from fragments of text. The word agile in the expression "agile sentiment analysis" is not just another buzzword; it is meant to denote that the sentiment analysis is to be carried out in a simple and efficient manner, even for the sake of non-perfect sentiment classification results.

It is a common practice that organizations keep track of the received media questions and their public relations (PR) responses and messages. We propose an approach that, besides merely storing and documenting past data records, advances the analysis of such data with agile sentiment analysis. The motivation for approach was to analyze past questions with respect to the sentiment and to identify actionable criteria that can be used to predict and at least in part control the sentiment of future questions. The hypothesis is that the more active role an organization plays, the more positive sentiment questions it receives in the long run.

The paper is structured as follows. In the next section we present a typical workflow for communicating between the media and organizations. We describe an approach that analyses the sentiment of received questions. The approach is evaluated on a selected case observed between a public fund and the media. We present the results of analysis and conclude the paper with summarizing the most important findings.

## **2 Method for agile sentiment analysis**

Organizations under severe media attention typically outsource their PR function to increase effectiveness and reduce the cost [L07]. Still, a lot of work has to be carried out by the internal employees. For example, official ingredients (i.e. contents) of the communications are within the responsibility of internal employees.

During the ordinary conduct of the work typical governmental organization receives several inquiries about their actions and decisions. The emotional context of such questions can vary from negative to neutral or positive. We present a method for storing and analyzing questions received and answers generated by an organization. The key ingredient of the presented method is agile sentiment analysis that is performed on the received questions. It is a common understanding that maintaining a good public reputation and handling media relations is of crucial importance for organizations. The

process of receiving questions and generating answers is depicted as a workflow in Figure 1.

The crucial part of the approach is encapsulated in the task for Analyzing text data. The task starts by standard text mining data preprocessing including removing stop words and lemmatization [FS06]. Next, for each question the probabilities of each of the three sentiments (negative, neutral, positive) were computed using Naïve Bayes Classifier for text classification. We trained the Naïve Bayes Classifier on the set of 345 preselected short questions in Slovene language containing negative, neutral and positive wording.

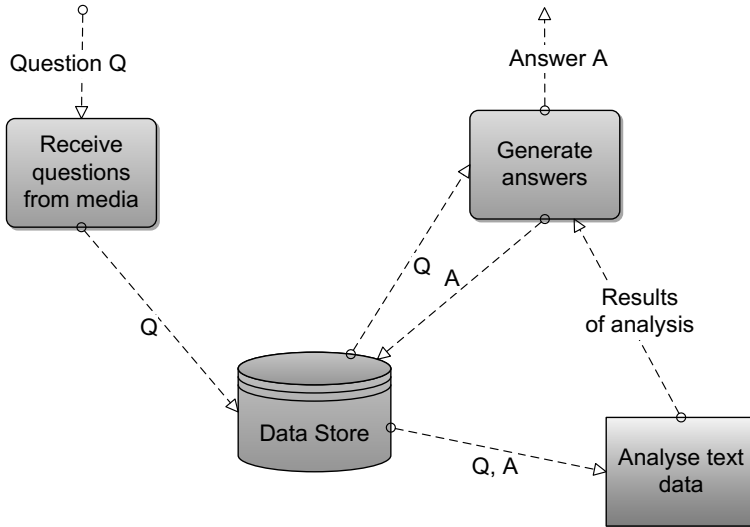


Figure 1: Workflow for receiving questions from media and generating answers

In Table 1 the words that are the most “negative”, “neutral” or “positive” are given with their computed probabilities according to the formula (2). Note that the formula (1) is used to compute probability of a word  $w_i$  given that the sentiment is positive. The constant  $\varepsilon$  is used as a smoothing constant in order to alleviate problems with probabilities close to 0 and 1 [G65] for the  $n$  class problem setting. In our experiment we took  $\varepsilon = 0.1$  and  $n = 3$ . The formula (2) is Bayesian reverse probability that is used to compute the probability of a positive sentiment given the word  $w_i$ . By the same token, probabilities for neutral and negative sentiment can be computed.

The final probabilities for the three sentiment classes were then combined into a single sentiment number from the interval  $[-3 .. 3]$ , where -3 is the most negative sentiment, 0 neutral and 3 most positive sentiment of a particular question.

$$p(w_i | \oplus) = \frac{N(w_i \cdot \oplus) + \varepsilon}{N(\oplus) + n \times \varepsilon} \quad (1)$$

$$p(\oplus | w_i) = \frac{p(\oplus) \times p(w_i | \oplus)}{p(w_i)} \quad (2)$$

word $w$	$p(\odot   w)$	$p(\ominus   w)$	$p(\otimes   w)$
advantage	0,50	0,31	0,19
efficient	0,54	0,01	0,45
kind	0,55	0,30	0,15
...			
blame	0,20	0,01	0,79
angry	0,19	0,00	0,80
reject	0,11	0,07	0,82
...			
saving	0,45	0,54	0,01
good	0,30	0,56	0,14
return	0,28	0,64	0,08

Table 1: Probabilities of the sentiments given the selected words.

Our decision to extend the number of possible sentiment values was motivated by the fact that we wanted to use such more detailed values in further analyses. The combined number was computed according to (3). Note that emoticons are used for positive, neutral and negative sentiments in this formula.

$$S = \text{ROUND} ( p(\odot) * 7 + p(\ominus) * 4 + p(\otimes) * 1 ) - 4 \quad (3)$$

As a result of the above procedure, each stored question had a corresponding sentiment value. Based on the determined sentiment value several additional analyses can be performed. The results of concrete analyses are presented in the next section.

### 3 Results on a selected case

In this section we demonstrate the proposed approach on data about communication between the media and a public organization that operates in the housing area and is responsible for financially supporting the Slovenian national housing programme. The Housing Fund of The Republic of Slovenia was founded in 1991 as one of the necessities of the Housing Law. In the last two decades the Fund's resources were primarily allocated for loans with financially pleasing terms to citizens and non-profit housing organizations. In addition, the Fund's financial incentive was used to increase the supply of newly constructed flats to the housing real estate market, as well as to encourage housing savings and granting subventions to young families for their first attempt to consolidate their housing status. Due to the delicate nature of the housing and financial business the Fund's activities received considerable media attention and were often criticized and disputed.

For practical demonstration of the approach we collected 298 journalists' questions and The Fund's answers in the period between October 2007 and November 2012. Besides the questions and answers we have within the selected period also identified 103 press releases, 41 explanations, and 8 press conferences issued and organized by The Fund.

Figure 2 shows average sentiment values for the received questions with respect to time. The average sentiment value of a question is -0.66. Negative sentiment gaps are clearly visible. In a few concrete cases we were able to map some external and internal events that resulted in such negative sentiment oscillations. For example, in period from July till October 2008 the Fund was under severe bombardment with questions about political orientation of its management, actual business policy orientation and capital expenditure issues. The interesting part is that the majority of the negative sentiment questions came from newspaper Finance, which is clearly visible from Figure 3.

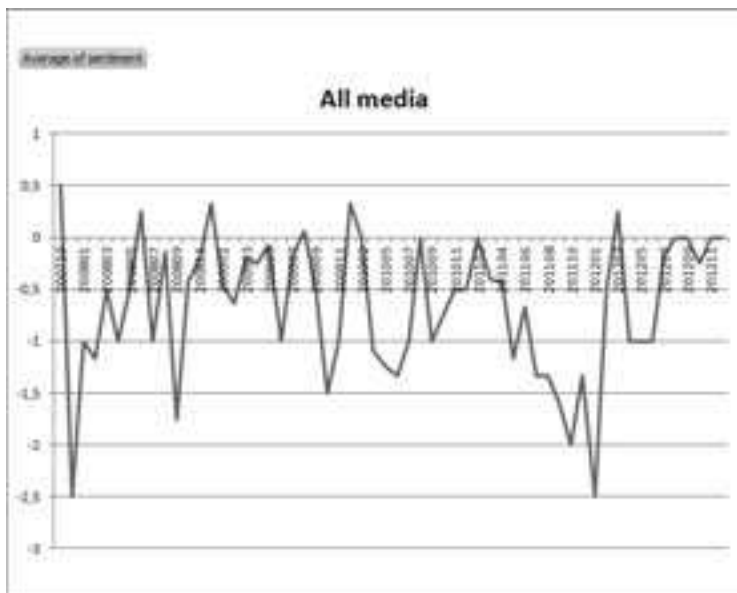


Figure 2: Average sentiment of the media questions received by The Housing Fund from October 2007 till November 2012. Number of all questions received is 298.

From Figure 3 it can be observed that Finance is the most critical media with respect to negative sentiment questions. However, a turnout in their publishing policy can be observed after the second half of 2012, as their questions had become more neutral.

Another negative peak in October 2009 can be attributed to the hesitant behavior of the responsible ministry to assign budget resources for housing subventions for the given year. The Fund was responsible for carrying out the corresponding float for housing subventions and received, probably unjustly, many critical negative sentiment questions. Interestingly enough, the leading role of a troublemaker was at that time played by another newspaper Dnevnik (see Figure 3). The quest continued in during the first months of 2010 with questions mostly related to the stall of the project for building new

housing dwelling in the capital city, where many citizens were interested to rent or buy a flat.

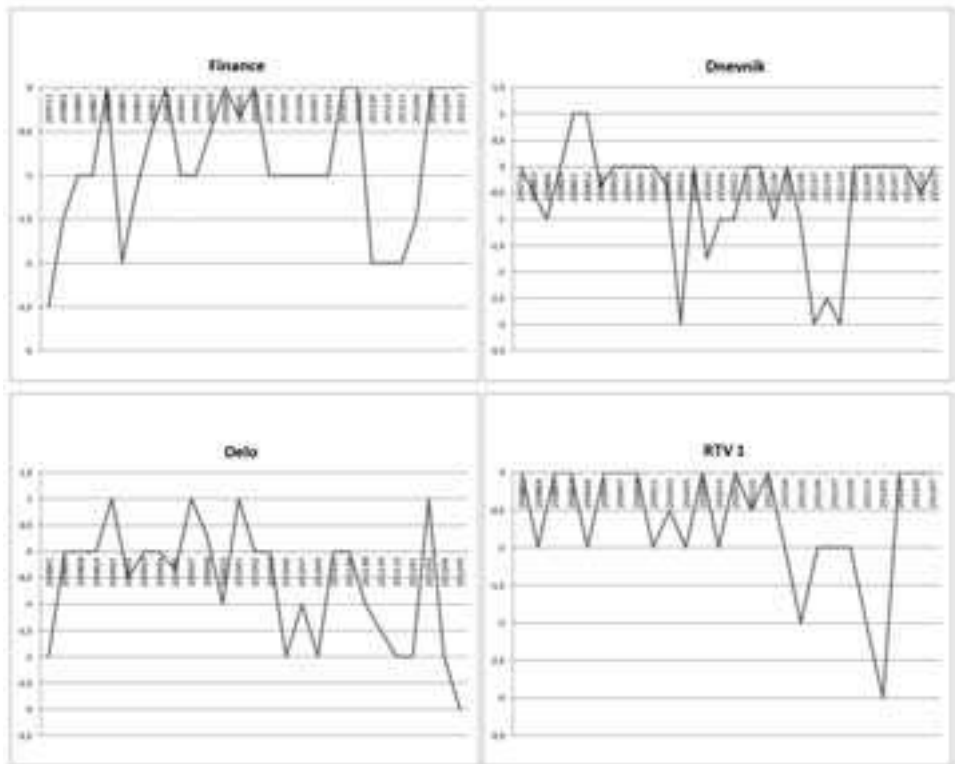


Figure 3: Average sentiment of the media questions from different media. The average sentiment value for Delo is -0.49, for Dnevnik -0.56, for RTV 1 -0.66, and for Finance -0.97. The sentiment for each question is estimated between -3 (negative) and 3 (positive).

Another negative gap in Figure 2 is evident in the second half of 2011. Here, the major media acted in coordinated fashion. The questions were again related to earmarked budget financial resources for housing subventions to young families, as well as to the stalled housing construction project.

Figure 3 shows the comparison of four largest media in terms of the average sentiment of their questions. The important lesson is that not all the media are sync with the negative or positive sentiment. The differences encountered are mostly due to different editorial policies of the media. What might be more alarming is the possibility that the sentiment of questions depends on the political alliance between pro-government media and the actual management structure. Systematic analysis and comparison of questions' sentiment between different media can reveal such discrepancies.

When analyzing the questions another interesting correlation was revealed. It is depicted in Figure 4. Each question received from the media is typically composed from several

minor questions. When performing the agile sentiment analysis we took into account the whole question. However, one interesting parameter can be also the number of actual sub-questions. In Figure 4 we can observe the correlation between the number of sub-questions and the average sentiment score. Namely, the higher the number of sub-questions, the more negative sentiment of the overall question is.

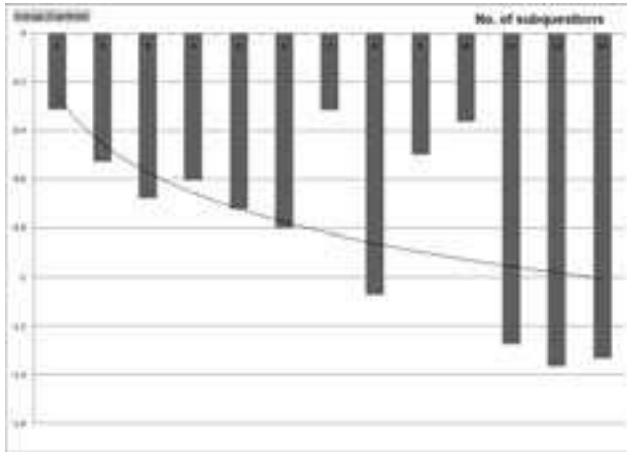


Figure 4: Correlation between the number of sub-questions in a question and the average sentiment of entire question. The trend line is calculated using logarithmic regression type.

## 4 Conclusions

In the paper we presented the approach to agile sentiment analysis of the media questions that is in regular use at the Housing Fund. Sentiment analysis helps officers react to questions more swiftly and in a coordinated fashion. The analysis also helps prioritizing the PR work. Note that the analysis revealed also rather surprising correlation between the question’s sentiment and the number of days needed for the answer shown in Figure 5. One of the consequences of this analysis was rather obvious decision of the fund’s management to take more active role in public relations by introducing more frequent and regular press conferences.

The perception of a sentiment of a particular media question is highly subjective. Sometimes a perfectly neutral question that touches a sensitive area can be perceived as rude or offensive. The system for agile sentiment analysis tries to avoid such subjective judgments by incorporating objective statistical properties in the estimation process.

However, the system itself has potential deficiencies. It fails to detect irony and sarcasm. Observed on a single sentence, the agile sentiment analysis might fail considerably. But so can sometimes a careless human mistakenly take sarcasm for kindness. However, as it was shown on numerous occasions, the system is statistically effective on larger datasets. Its main advantage is to automatize the process that would normally take a considerable

amount of time and produce highly subjective results at the end. For example, many authors successfully implemented agile sentiment analysis on Twitter [BR10].

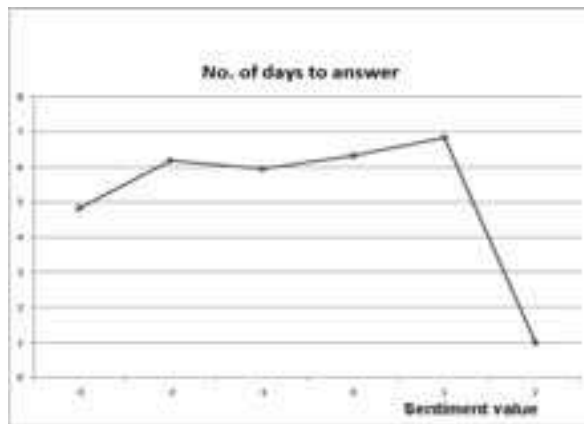


Figure 5: The surprising finding: all the answers to questions were generated within the legally binding period of 8 working days, while the answers to the questions with the most positive sentiment (2) were generated within a single day.

A typical journalist question should be neutral. However, in the modern world, where the best news that sell are bad news, the questions sometimes guide the answering person into “muddy” terrain where the answers can be understood as the journalist’s story requires and not as they were meant to be. Here, the assumption was that the respondent’s intentions were righteous and without any hidden agendas. However, there are cases when the journalists discover malicious behavior and conduct of the questioned person; as for such cases, the “negative sentiment” questions seem to be more than justified. Note that the presented system was not designed to distinguish between these two situations.

## References

- [BBC13] BBC News Business: Cyprus bailout: Dijsselbloem remarks alarm markets, BBC News Business, 25 March 2013, accessed at <http://www.bbc.co.uk/news/business-21920574>.
- [BR10] Batra, S.; Rao, D.: Entity Based Sentiment Analysis on Twitter. *Science* (2010), Volume: 9, Issue: 4, Pages: 1-12.
- [CRI12] Capgemini, Rand Europe, IDC, Sogeti and Dti: eGovernment Benchmark Framework 2012-2015. Method paper, July 2012.
- [FS06] Feldman, R.; Sanger, J.: *The Text Mining Handbook*. New York: Cambridge University Press, 2006.
- [G65] Good, I. J.: *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*, Cambridge, Mass., M.I.T. Press, 1965.
- [L07] Lietz, G.: Shared services: e-government and economies of scale. In Zehner (Ed.) *E-Government Guide Germany, Strategies, solutions and efficiency*, Fraunhofer IRB Verlag, Stuttgart, 2007.
- [L10] Liu, B.: Sentiment Analysis and Subjectivity. In *Handbook of Natural Language Processing* (eds.: Indurkha, Damerau), 2010.