

Towards Sentiment Analysis of Financial Texts in Croatian

Željko Agić*, Nikola Ljubešić**, Marko Tadić**

*Department of Information Sciences

**Department of Linguistics

Faculty of Humanities and Social Sciences, University of Zagreb

Ivana Lučića 3, 10000 Zagreb, Croatia

{zeljko.agic, nljubesi, marko.tadic}@ffzg.hr

Abstract

The paper presents results of an experiment dealing with sentiment analysis of Croatian text from the domain of finance. The goal of the experiment was to design a system model for automatic detection of general sentiment and polarity phrases in these texts. We have assembled a document collection from web sources writing on the financial market in Croatia and manually annotated articles from a subset of that collection for general sentiment. Additionally, we have manually annotated a number of these articles for phrases encoding positive or negative sentiment within a text. In the paper, we provide an analysis of the compiled resources. We show a statistically significant correspondence (1) between the overall market trend on the Zagreb Stock Exchange and the number of positively and negatively accented articles within periods of trend and (2) between the general sentiment of articles and the number of polarity phrases within those articles. We use this analysis as an input for designing a rule-based local grammar system for automatic detection of polarity phrases and evaluate it on held out data. The system achieves F1-scores of 0.61 (P: 0.94, R: 0.45) and 0.63 (P: 0.97, R: 0.47) on positive and negative polarity phrases.

1. Introduction

This paper describes the first in what is hopefully to become a series of experiments dealing with setting up a general framework for sentiment analysis of financial texts written in Croatian. Over the period of last three years, we have seen financial markets and stock exchanges of the world (the Zagreb Stock Exchange included) entering media mainstream due to the global impact of the financial crisis. From a natural language processing point of view, such an increase in the amount of domain-specific texts available online provokes interest and raises different questions. Some of these questions might look as follows:

- Is there any correlation between trade figures on the stock exchange and its online media coverage and is it measurable?
- Do online newspaper articles, blogs and forum discussions reflect opinions on the current and future trends for the stock exchange?
- Are there any specific events occurring on the stock exchange influenced by these opinions being disseminated to the general public?
- Could we predict future trends for elements of the stock exchange by automatically analyzing sentiment from various sources?

We have set a goal of providing answers to these thought-provoking questions, even if they might appear to be somewhat broad, as an underlying motivation for this experiment and our future research.

In terms of published experiments related to sentiment analysis in general and analysis of financial text sentiment specifically, that follow the lines of the questions we presented here, substantial work was done for English and some for German language. We find the experiments presented in (Ahmad et al., 2005; Ahmad et al., 2006a; Ahmad et al., 2006b; Daly et al., 2009; Remus et al., 2009)

to be of special importance for our own research. This series of papers deals with proving a correlation between the behaviour of a number of stock exchange indices and distributional properties of newspaper texts dealing with corresponding financial markets. Afterwards, they proceed with developing system prototypes for automatic analysis of these properties and correlations. The work presented in these papers is still ongoing. Experiments presented by (Devitt and Ahmad, 2007; Devitt and Ahmad, 2008) also follow similar guidelines. We consider the former to be particularly interesting, as it presents an algorithm which annotates polarity phrases, i.e. phrases encoding positive or negative sentiment, to unseen text by utilizing information on word polarity encoded in SentiWordNet (Esuli et al., 2006). Furthermore, (Engle and Ng, 1993) and (Smith, 1995) are also somewhat interesting for our research, as they deal with analyzing the impact of news on market index volatility and properties of domain texts encoding a certain sentiment, respectively. We also consulted (Turney, 2002), which presents an unsupervised algorithm for binary classification of reviews from four different domains by calculating the so-called semantic orientation of the phrases within texts.

The specific experiment presented in this paper deals with assembling and analyzing basic resources needed for data-driven sentiment analysis in Croatian financial texts, i.e. the manually annotated gold-standard text collection annotated with sentiment. The gold-standard collection consists of articles manually annotated for general sentiment (positive, negative or neutral) and polarity phrases (positive or negative) within text. These two levels of manual annotation were analyzed with respect to each other and the trends exhibited by the market index of the Zagreb Stock Exchange (<http://www.zse.hr>). Additionally, the usefulness of the approach is illustrated by using results of this analysis as guidelines for

designing, developing and evaluating a rule-based local regular grammar module for detection and binary classification of polarity phrases in Croatian texts from the financial domain.

The following sections of the paper describe various properties of the language resources created for purposes of the experiment, the design and development of the local grammar system for polarity phrase detection and its performance on Croatian texts from the financial domain. The paper concludes with an insight on perspectives for future research.

2. The experiment

In this section, we provide an analysis of the language resources we collected for purposes of the experiment and present an evaluation of the rule-based polarity term detection module we developed.

2.1 Document collection

Our collection of financial texts has been compiled from two distinct information sources: (1) online newspapers specializing in finance and (2) a large forum discussing the Zagreb Stock Exchange. Namely, we collected a total of 413.153 articles specifically dealing with financial markets from *business.hr* (<http://www.business.hr>), *limun.hr* (<http://www.limun.hr>) and *poslovni.hr* (<http://www.poslovni.hr>) online magazines and articles of various business-related topics from the Croatian subsection of the *seebiz.eu* finance portal targeting the South-Eastern Europe region (<http://www.seebiz.eu>). We also collected more than 200.000 forum posts from the *poslovni.hr* stock market forums (<http://www.poslovni.hr/forum>) discussing the Zagreb Stock Exchange. Articles were extracted for the interval between October 2005 and October 2009, but only articles between 2006-09-01 and 2009-09-01 are used in this research. This makes up for a total of 342.703 articles. We excluded the earlier periods, being that they were less turbulent for the stock exchange and that online finance portals and forums were just kick-starting back then, having fewer texts with an unbalanced daily distribution. Forum posts are currently excluded from this line of research due to specifics of their language in comparison with newspaper articles. Article statistics are given in table 1.

Name	Articles	Used here	Tokens	Categories
<i>business.hr</i>	13.248	12.264	3.094.635	yes
<i>limun.hr</i>	254.062	216.426	29.329.045	no
<i>poslovni.hr</i>	123.000	98.623	15.509.429	no
<i>seebiz.eu</i>	22.843	15.390	6.256.745	no

Table 1 Properties of the document collection

In order to analyze the article sentiment, we chose two long periods of evident trend on basis of data acquired for the Zagreb Stock Exchange index (CROBEX). The positive or bullish period was between 2007-01-02 and 2007-05-21 and the negative or bearish period dated from

2008-01-07 to 2008-04-16. From these two periods of trend, we selected articles at random from the *business.hr* collection to be manually annotated for sentiment by two annotators. Each annotator was assigned 600 articles with 100 of them overlapping, making for an overall count of 1100 different articles or approximately 250.000 tokens. The annotators were instructed to annotate the articles for overall sentiment using three basic categories: positive (+), negative (-) and neutral (0). They were also given a subtask of annotating specific expressions or language constructs they considered to be indicating polarity in 30 articles from the overlapping section. The polarity terms were annotated using only positive (+) and negative (-) flags.

By now, the two human annotators have assigned the overall sentiment tags to 534 articles (ca 125.000 tokens), with 44 of them overlapping. A total of 18 out of the 44 overlapping articles contained annotations of specific polarity terms. The annotators agreed in general sentiment on 35 and disagreed on 9 articles, making for an observed agreement of 0.8977. The 18 articles annotated with polarity terms contained 259 terms with annotations, 189 of them overlapping and 70 not overlapping, with an observed agreement of 0.8148. None of these overlapping terms were of distinct sentiment. Linearly weighted kappa coefficients for general article sentiment agreement and polarity term agreement between the annotators were calculated to be 0.7762 and 0.7174 respectively. Both are generally considered to be good agreement scores.

	Positive sent.	Negative sent.	Neutral sent.
Bull market	143	45	69
Bear market	67	96	70

Table 2 General article sentiment over periods of trend

	Positive sent.	Negative sent.	Neutral sent.
Positive terms	112	71	16
Negative terms	23	170	10

Table 3 Polarity terms over general article sentiments

	Positive term	Negative term
Positive term	137	49
Negative term	47	140

Table 4 Distribution of neighbouring polarity terms

Tables 2 and 3 provide an insight on the relationship between overall article sentiments over periods of obvious trend and between overall article sentiment and polarity terms contained within those articles, respectively. Relations between positive and negative polarity articles and periods of trend (table 2) and between positive and negative polarity articles and terms (table 3) were used to calculate statistical significance of variable independence using the Fisher's exact test (cf. Agresti, 1992). The test itself, even though used primarily on contingency tables with lower event counts, is exact and was calculable for

this data. The obtained p-values were found to indicate statistical significance of the results, their values being 3.355e-11 and 2.2e-16, respectively.

Table 4 illustrates how positive and negative polarity terms detected by human annotators are spread in the course of the articles. It clearly shows that polarity terms are more often followed (or neighbored) by other terms of the same polarity than by terms of distinct polarity. The p-value of the Fisher's test on this table was less than 2.2e-16. This supported our intuition that news article writers would tend to write their articles in blocks or paragraphs sharing the same sentiment.

2.2 Polarity phrase detection module

During collection of the frequency list of all manually annotated polarity phrases we noticed that their structure is predominantly predictable and regular. Therefore, we decided to use the series of local (regular) grammars applied in cascades to create the module for automatic detection of these phrases. The system was implemented using the NooJ development environment (Silberztein, 2004; Silberztein, 2006) and the Croatian lexical resources for NooJ (Bekavac et al., 2007a).

The module itself consists of two disjoint sub-modules: the sub-module for detecting positive polarity phrases and the sub-module for detecting the negative ones. Such grammar design implicitly enables the classification of polarity phrases alongside their detection. The local grammars are implemented as cascades of finite state transducers, giving XML-annotated output denoting positive and negative polarity. This is illustrated by figure 1 and a sample NooJ local grammar for detecting certain positive polarity terms is illustrated in figure 2.

```
<ph p="pos" w="">Europske dionice u četvrtak su ojačale</ph>
  poduprte <ph p="pos" w="">ohrabrujućim rezultatima
poslovanja</ph> kompanija i većim kupnjama dionica koje su pojeftinile
zbog ovotjednog <ph p="neg" w="">oštrog pada cijena dionica</ph>.
```

Figure 1 Transducer output

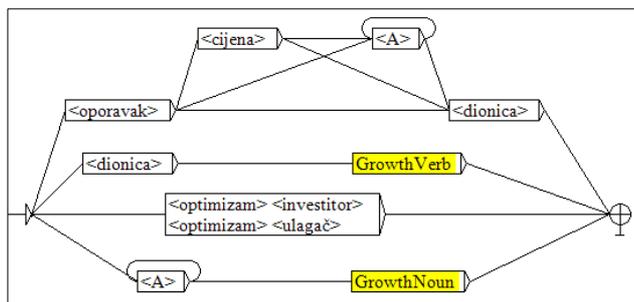


Figure 2 Local grammar sample from NooJ

In figure 1, we can see that the XML tag *ph* denotes phrase carrying sentiment information, while its attributes *p* and *w* encode polarity (with values being either *pos* or *neg*) and the weight of the expression (currently left empty, but reserved for planned future work). The local grammar illustrated by figure 2 detects expressions such as e.g. *oporavak cijena europskih dionica* (en. *European stock price recovery*) and *dionice uzletjele dodatnih 5*

posto (en. *stocks flew up an additional 5 percent*).

The module was evaluated on the data held out from the gold standard and the results are presented in table 5. An additional assessment of module precision was conducted manually on the entire business.hr article collection, but the scores achieved this way were virtually identical to the ones presented in the table.

Sub-module	Precision	Recall	F1-measure
Positive	0.94	0.45	0.61
Negative	0.97	0.47	0.63

Table 5 Sub-module accuracy

Precision of the module on both positive and negative polarity phrases is very high. Combined with significantly lower recall scores, it yields a satisfactory F1-measure of 0.61 on positive and 0.63 on negative polarity phrases. Lower recall scores are a direct consequence of a rather conservative design of local regular grammars that we imposed during module development. Lowering the precision by implementing additional relaxed rules would undoubtedly result in raising system recall and detecting more polarity phrases.

3. Conclusions and future work

We have presented the results of a preliminary experiment in sentiment analysis of Croatian financial texts. We assembled a collection of more than 400.000 documents in Croatian from the domain of finance. Selecting from this collection, we also created a small gold-standard corpus of newspaper articles and annotated it manually for general sentiment and polarity phrases within text. We have shown that the overall market trend, as defined by the index of the Zagreb Stock Exchange, correlates with the distribution of general sentiment within articles taken from the corresponding periods of trend. Furthermore, we have also shown that general sentiment of the articles is in turn governed by the polarity phrases contained within these articles, i.e. that articles carrying positive overall sentiment will usually contain more positive polarity phrases than the phrases with negative polarity, making it possible to calculate overall polarity of unseen articles by counting polarity phrases detected within them. We have also developed a prototype module for automatic detection and classification of polarity phrases in unseen Croatian texts from the financial domain by using cascade of local (regular) grammars. The module was shown to perform with high precision and satisfactory recall on unseen texts.

The experiment presented here was preliminary in a sense that its purpose was to detect whether for a highly inflectional and free order language as Croatian is (as the rest of Slavic languages also), this kind of analysis is feasible. The results have encouraged us to lay out here possible future research plans that are quite extensive. They will probably be spread along these lines:

- Expanding and enriching the manually annotated gold standard collection, combined with additional fine-tuning and possibly re-evaluation the polarity phrase detector/classifier. We consider the task of

raising module recall to be particularly important.

- Evaluating the detection and classification module in terms of reflecting its results back to the stock market index. Namely, we should investigate the output of the module with regards to stats given in tables 2 and 3 to see if the noise it introduces by low recall disrupts the regularities observed when correlating the trends of the stock market and the manual annotation.
- Shifting our sentiment analysis framework to the task of intra-day sentiment detection and trend correlation. This task would have been reachable even for this experiment, if the intra-day statistics for the Zagreb Stock Exchange index were made freely available at the time of conducting this analysis.
- Our rule-based polarity phrase classifier is currently being expanded to add certain numeric weights to detected phrases instead of only classifying them. This way we could investigate the influence of phrase weight on overall article sentiment. This builds on an intuition that phrases do indeed carry different degrees of severity. For example, *a steep decrease* is obviously a more substantial decrease than *a mild decrease* and we argue that such phenomena should be accounted for in the system.
- We would like to combine our system with the named entity classifier for Croatian (Bekavac and Tadić, 2007b; Bekavac et al., 2009) in order to investigate trends for specific stocks listed on the Zagreb Stock Exchange, rather than inspecting the combined index. We believe that this course of action might provide additional insight on the relation between e.g. certain shared properties of stock trading volumes and sentiment distributions.

4. Acknowledgements

The research within the project Accurat leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013), grant agreement no 248347. This work was also supported by the Ministry of Science, Education and Sports, Republic of Croatia, under the grants No. 130-1300646-1776, 130-1301679-1380 and 130-1300646-0645.

5. References

Agresti, A. (1992). A Survey of Exact Inference for Contingency Tables. *Statistical Science* 7 (1), pp. 131–153.

Ahmad, K., Gillam, L., Cheng, D. (2005). Society Grids. *Proceedings of the UK e-Science All Hands Meeting 2005*. 18-21 September, Nottingham UK. Swindon: EPSRC Sept 2005. pp 923-930.

Ahmad, K., Lee Gillam and Cheng, D. (2006). Sentiments on a Grid: Analysis of Streaming News and Views. *Proceedings of the 5th International Conference on Language Resources and Evaluation*.

Ahmad, K., Cheng, D., Almas, Y. (2006). Multi-lingual Sentiment Analysis of Financial News Streams. *First International Conference on Grids in Finance*. Palermo,

Italy, 3-4 February 2006, , International School for Advanced Studies, Trieste, Italy, 2006.

Bekavac, B., Vučković K., Tadić, M. (2007). Croatian Resources for NooJ. *2007 NooJ Conference Book of Abstracts*. Barcelona, 2007.

Bekavac, B., Tadić, M. (2007). Implementation of Croatian NERC system. *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing 2007, Special Theme: Information Extraction and Enabling Technologies*. Association for Computational Linguistics, Prague 2007, pp. 11-18.

Bekavac, B., Agić, Ž., Tadić, M. (2009). Interacting Croatian NERC System and INTEX/NooJ Environment. *Proceedings of the 2008 International NooJ Conference*. Cambridge University Press, Cambridge 2009, in press.

Daly, N., Kearney, C., Ahmad, K. (2009). Correlating market movements with consumer confidence and sentiments: a longitudinal study. *Text Mining Services 2009*. Leipzig, Germany, 23 March 2009, Leipziger Beitrage zur Informatik, 2009, pp. 169-180.

Devitt, A. and Ahmad, K. (2007). Sentiment Polarity Identification in Financial News: A Cohesion-based Approach. *Annual Meeting of the Association of Computational Linguistics (ACL 2007)*. Prague, Czech Republic, 25-27 June, 2007.

Devitt, A., Ahmad, K. (2008). Sentiment Analysis and the Use of Extrinsic Datasets in Evaluation. *Proceedings of the 6th International Conference on Language Resources and Evaluation*. Marrakech, Morocco, 28-30 May, 2008.

Engle, R., Ng, V. (1993). Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, Vol. 48, No. 5 (Dec 1993), pp. 1749-1778.

Esuli, A., Sebastiani, F. (2006). SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. *In Proceedings of LREC-06, 5th Conference on Language Resources and Evaluation*. Genova, Italy, 2006, pp. 417-422.

Remus, R., Heyer, G., Ahmad, K. (2009). Sentiment in German language news and blogs, and the DAX. *Text Mining Services 2009*. Leipzig, Germany, 23 March 2009, Leipziger Beitrage zur Informatik, 2009, pp. 149-158.

Smith, G. (1995). How High Can a Dead Cat Bounce? Metaphor and the Hong Kong Stock Market. *Hong Kong Papers in Linguistics and Language Teaching*, volume 18, 1995, pp 43-57.

Silberztein, M. (2004). NooJ: an Object-Oriented Approach. *INTEX pour la Linguistique et le Traitement Automatique des Langues*. Cahiers de la MSH Ledoux. Presses Universitaires de Franche-Comté, pp. 359-369.

Silberztein, M. (2006). NooJ Manual. Available at URL <http://www.nooj4nlp.net/NooJ%20Manual.pdf>.

Turney, P. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 2002, pp. 417-424.

The Zagreb Stock Exchange website. Available at URL <http://www.zse.hr> (last accessed 2010-03-15).