Piotr Młodzianowski

Warsaw University of Technology, Faculty of Management, Chair of Finance and Financial Systems, piotr.mlodzianowski@pw.edu.pl

A Study of the Influence of Online Information on the Changes in the Warsaw Stock Exchange Indexes

Abstract: The article presents the results of a study on the influence of online information originating from financial websites on changes in the Warsaw Stock Exchange indexes. The first part is theoretical. It describes the issue of text mining and sentiment analysis and their use in the text analysis process. The next part of the article describes the characteristics of the study. A selection was made of Polish financial websites that may trigger reactions from investors on the Warsaw Stock Exchange. Words occurring on the analysed websites were selected and put into classes. Then the relation between changes in WSE indexes and the frequency of appearance of individual words within the classes was analysed. The last part of the article presents the study results, discusses the possibilities of using them and indicates further areas for research.

Keywords: news, Warsaw Stock Exchange, text mining, sentiment analysis

JEL: G12, G14

1. Introduction

Collecting, processing and using information is an essential part of the development of civilisation. Today the internet and its resources are the fastest way of acquiring information. Huge amounts of unstructured data, such as commentaries, photos, reports, contracts, offers, regulations, etc. are kept on companies' and organisations' servers. Although the unit cost of storing data has been falling, an increasing challenge has arisen of monitoring these vast resources and at the same time separating information items that are important from those that are not, as well as the ones which are true from those which are false.

Electronic media have become the sources of information as they have offered increasingly cheap and easy access to information. Information is created and shared by users on news portals and the entire process has become an element of social behaviour (Ling, 2012). Information is not only created by people, as its increasing share is the result of automated work, such as analyses of geological data in search of potential earthquakes or transaction systems on financial markets, operating in order to make a specific investment decision. Taking into account the rate of growth of available information, one could speak of information explosion (Hilbert, 2012: 8–12). It is a dynamic process consisting of the increase of the amount of available information, in particular due to:

1) increased rate of producing new information;
2) easiness of reproducing and transmitting data on the internet;
3) increased number of available incoming information channels;
4) large amounts of collected historical data;
5) no method of processing or comparing various types of information, often conflicting and imprecise, and duplication of available information.

The advantages of information explosion include better and cheaper access to information, faster publication and the creation of new professions and related jobs in information processing. Alongside the advantages there are also threats that result from the discussed phenomenon, including: increased costs of information processing, difficulties in separating true information from false, no possibility of "being forgotten" on the internet and working time losses due to increasing numbers of emails, phone calls and information items reaching employees (Dutta, 2013: 48–130). The purpose of this article was formulated in the context of the above-mentioned advantages and threats, namely the assessment of the possibilities to forecast changes on the Warsaw Stock Exchange (WSE) based on an in-depth analysis of information published on the internet. Therefore the primary goal may be formulated as:

PG: Defining the relation between information originating from websites and changes in WSE indexes.

Pursuing the above-mentioned primary goal, a study was developed and carried out that comprises the following targets:

T1: identification of Polish financial websites which may reflect customers' reactions to WSE changes;

T2: identification and selection of keywords for analysis and their division into classes;

T3: assessing the impact of website occurrence of selected word classes on stock exchange indexes.

Referring to the defined goal, it should be noted that most resources on the internet have the form of text documents lacking a defined structure, which hinders their automatic processing. The exploration of this gigantic repository is facilitated by smart text mining systems and sentiment analysis which make it possible to search, classify, summarise and interpret information. This article presents the possibilities offered by such analyses and the results of preliminary research on data published on websites linked to WSE customers.

2. Text mining and sentiment analysis

Text mining is a method of utilising unstructured text documents. First references to text mining can be found in a 1958 article by H.P. Luhn on automatic creation of abstracts, which describes the role of keywords in the source text (Luhn, 1958: 159–165). The assumptions for text mining were developed in 1960, with the construction of the first computer systems processing unstructured text. Further development of tools for explorative text analysis came about in the 1990s, with the birth of new branches of science: natural language processing (NLP) and artificial intelligence (AI), on which contemporary text mining is based. Research on methods of exploring unstructured data seems to be much needed as it helps to save time and money that would otherwise have to be spent on reading and potential exploration of the huge repository of text documents by man.

Text mining is increasingly often enhanced by sentiment analysis. It is a method of analysing qualitative data for emotionally-charged words. Sentiment analysis is based on two assumptions. First, some words express emotions. Second, there are words whose utterance may evoke emotions (Pang, Lee, 2008: 1–135). Therefore sentiment analysis indicates the emotional state of the author of the expression, and on the other hand defines the emotional effect that a given expression may have. The term 'sentiment analysis' in this sense was introduced by Das and Chen (2001: 43) and Tong (2001: 1–6).

Analysis of opinion (Pang, Lee, 2008: 1–135), an example of which is the sentiment analysis, uses solutions developed in the field of natural language processing (Nasukawa, Yi, 2003: 70–77). Its practical application was accompanied by fast development of dictionaries for analysing statements and documents (Nielsen, 2011: 93–98). On the one hand we have thematic dictionaries that classify expressions according to their

subjects, and on the other hand we have seen the development of various dictionaries that make it possible to identify words and statements that express or evoke emotions. These dictionaries allow for simple classifications (positive-negative) as well as more complex classifications (anxiety-glory-aggression-sadness-love). There have also appeared mixed dictionaries that combine both ideas. An example of such a tool is the dictionary by Loughran and McDonald (2011: 35–65), which classifies statements related to economics and finance according to the emotional charge included in them.

One of the first people who noticed the possibility of using the presented tools to analyse financial markets was Lupiani-Ruiz. He built a financial news search engine (Lupiani-Ruiz et al., 2011: 15565–15572). It was limited to searching for numerical values in the text. The possibilities to use financial news in forecasting the direction of stock index movements were intensively researched from the beginning of the 21st century, with varying results (Hagenau, Liebmann, Neumann, 2013: 685–697; Mittermayer, 2004: 10; Schumaker, Chen, 2009: 1–19; Tetlock, Saar-Tsechansky, Macskassy, 2008: 1437–1467). Research was also conducted on the FX market (Peramunetilleke, Wong, 2002: 131–139; Nassirtoussi et al., 2015: 306–324). The studies looked for relationships between pieces of information, news items and changes on the market.

The most popular method is the so called "bag of words" approach. It treats the frequency of occurrence of particular words in the document as attributes, and then searches for relations between them and changes on the market. The place and sequence of words is disregarded. The multidimensionality of the space of attributes created this way poses a significant problem. This is because typical texts contain between several thousand and tens of thousands of words. Therefore methods are sought to choose words or groups of words that are semantically the most significant for a given set of documents or words are initially grouped into classes. The classes represent words with similar meaning or ones expressing similar emotions. The method also has disadvantages. Words written in the same way may have different meanings, in particular when the diacritical marks that are elements of the letters ą, ć, ę, ł, ń, ó, ś, ż and ź are removed. A word's meaning may also change due to the preceding words or depending on the context.

The results of research that uses the above-mentioned elements to determine the possibility of forecasting the WSE participants' reactions based on text mining and sentiment analysis of selected words and word classes are presented below.

3. Research characteristics

Identification of relevant and required data is one of the most important tasks in the analysis. The roles of explanatory variables (forecasting variables) and dependent variables (forecast variables) should be determined. The sources of press informa-

tion should be defined as well as whether it would be used to model stock indexes or quotations of individual companies. For the purposes of this study it was decided that information found on the most popular websites focusing on "Business, Finance and Law" according to a January 2015 survey by Megapanel PBI/Gemius be used (Wirtualnemedia, 2017). It contains a list of 20 most popular websites according to user numbers. In order to optimise the research process and make it less time-consuming, a research sample was selected that consisted of 6 websites, servicing 68% of the total number of users. They included: wp.pl (Money.pl), onet.pl, gazetaprawna.pl, bankier.pl, gazeta.pl and interia.pl.

The research covered information from the homepages and the first linked pages of those websites. The content of the pages was downloaded, but user comments under the articles were rejected to ensure the objectivity of the research. Another stage was to decompose the downloaded content into single words. Then the frequency of appearance of particular words was counted.

The analysis of selected websites was carried out every day at 8:50 am, before trading at the WSE started, and at 5:30 pm, after the trading ended. The analysis lasted about 5 minutes. A decision on the direction of stock index changes was made based on the results. It should be noted that all information available at 8:50 am was taken into account, regardless of the publication time. The analyses at 5:30 pm were used to determine the correctness of keywords choice. The observations were conducted between 21 Sept 2015 and 31 Oct 2016 and covered 280 trading days.

Original software implemented within MS Excel was used to convert the stream of characters into individual words. Then keywords were looked for in the aggregate word database. Thanks to the adopted form of identification, there was no need to use the basic form of the word. The searched words were conjugated and declined. This made it possible to look for the same words which differed only in the grammatical form, and then count and divide them into two classes: positive and negative. Due to the fact that predefined content was sought, issues related to proper interpretation of punctuation marks and clarification of the meaning of words spelt in the same way (e.g. the Polish word 'piła' may mean a person who has been drinking, a ball or a saw) were disregarded. The disadvantage of this approach is that it does not take into account the meaning of the word depending on its context.

The next stage of the research was to build the occurrence matrix which transformed the set of searched and classified words into a quantitative format. The matrix's row is made up of words appearing on a given day on news portals. The column is made up of words from the positive (Kp) and negative (Kn) classes. The occurrence matrix cell can be defined as:

$$Occurence\ matrix_{i,j} = f(the\ number\ of\ occurrences\ of\ word\ i\ on\ day\ j). \quad (1)$$

The last column of the occurrence matrix is the assessment of the Information Environment Sentiment before the start of trading (NOI_j), which is the difference between the frequency of occurrence of positive and negative class words. It is calculated in the following way:

$$NOI_j = Kp_j - Kn_j. \tag{2}$$

If:
$NOI_j > 0$, forecast direction of index change on day j is up,
$NOI_j < 0$, forecast direction of index change on day j is down,
$NOI_j = 0$, lack of forecast on day j,
where:
NOI_j – Information Environment Sentiment before start of trading on day j,
Kp_j – number of positive class words on day j,
Kn_j – number of negative class words on day j.

NOI is compared to the change in the stock index value that occurred on the same day.

The quantitative value of the change, calculated on the analysis day, is attributed to the forecast direction of index change by combining text data with time series. One should consider what time is needed for the state of knowledge at the time of analysis to be reflected in the values of the indexes, or how long it will take for the information to become incorporated into the price. Considering the fact that the analysed information is freely available to any user, the time of its "absorption" by the market should be close to zero. The research assumed two analysis times: 9:00 am – the value of analysed indexes as the trading opens (period soon after the analysis) and 5:00 pm – the value of indexes as the trading closes. The selection of times results from data accessibility.

The information impact was measured by the value of index change as expressed in points. If the forecast direction is in line with the index change direction, then the value of the change is treated as profit, and otherwise as loss. The rate of index change is calculated at 9:00 am (*IndexChangeOpening*) and 5:00 pm (*IndexChangeClosing*) in the following way:

$$IndexChangeOpening_{w,j} = OpeningPrice_{w,j} - ClosingPrice_{w,j-1}, \tag{3}$$

$$IndexChangeClosing_{w,j} = ClosingPrice_{w,j} - OpeningPrice_{w,j}, \tag{4}$$

where:
w – index name,
j – survey date.

If news was published after the trading session, its incorporation by the investors was possible only when the next day's trading started.

Similarly to the occurrence matrix construction, the collection of searched and classified words at 5:30 pm on a given day had to be transformed into quantitative data. The occurrence matrix cell was defined as:

$$Occurrence\ matrix'_{i',j'} = f'(number\ of\ word\ i'\ occurrences\ on\ day\ j'). \qquad (5)$$

The last column of the occurrence matrix is the assessment of the Information Environment Sentiment after the end of trading (NOI'_j), which is the difference between the frequency of occurrence of positive and negative class words. It is calculated in the following way:

$$NOI'_j = Kp'_j - Kn'_j. \qquad (6)$$

If:

$$NOI'_j > 0 - \text{direction of index change is up,}$$

$$NOI'_j < 0 - \text{direction of index change is down,}$$

$$NOI'_j = 0 - \text{no index change,}$$

where:
NOI'_j – Information Environment Sentiment after end of trading on day j,
Kp'_j – number of positive class words on day j after end of trading,
Kn'_j – number of negative class words on day j after end of trading.

The NOI' value is compared to the rate of change of stock indexes at the end of trading ($IndexChangeClosing'_{w,j}$) which occurred on the same day. It is calculated in the following way:

$$IndexChangeClosing'_{w,j} = ClosingPrice_{w,j} - ClosingPrice_{w,j-1}, \qquad (7)$$

where:
w – index name,
j – survey date.

If NOI after the close of trading is identical to the direction of index movement, the value of the change on this day is qualified as value that was successfully forecast using the selected words and created classes. If the value NOI'_j takes a different direction of change than the stock exchange index, the value of the index change is classified as value that was not successfully forecast. This makes it possible to determine whether the selected words describe the changes in the stock indexes to a sufficiently high degree (higher than the toss of a coin = 50%), and whether they could be used to forecast index change.

Based on the analyses and comparisons, conclusions were drawn on the possibility of using online information from websites to forecast the movement of stock indexes. In order to attain the primary goal and the targets, the analysis focused on finding answers to the following questions:

Q1: What financial websites are the most popular among Polish stock investors? Finding the answer to research question Q1 will make it possible to attain target T1.

Q2: Do the selected positive and negative words describe changes to stock indexes? Finding the answer to research question Q2 will make it possible to attain target T2.

Q3: To what extent do the selected word classes correspond to the changes in the direction of a stock index? Finding the answer to research question Q3 will make it possible to attain target T3.

4. Research results

The analysis of websites focusing on "Business, Finance and Law" made it possible to identify the most popular websites among Polish stock investors. It showed that six websites included in the research attracted 68% of the total number of users (Figure 1).
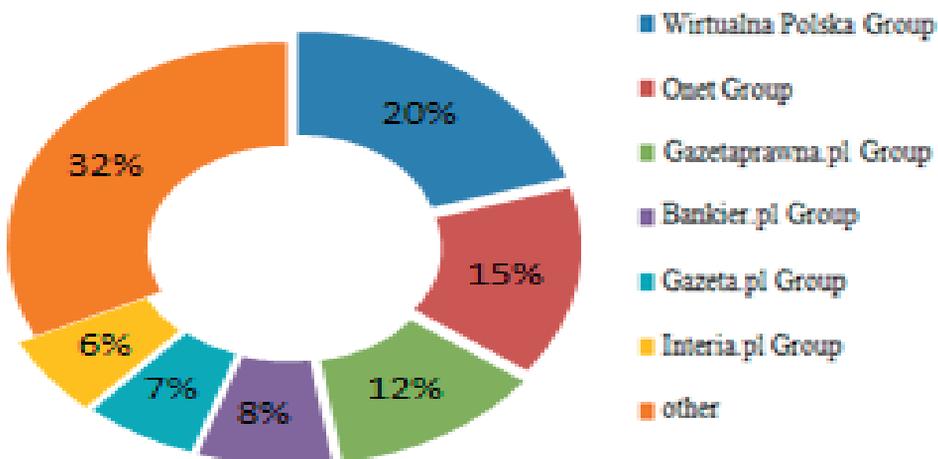

Figure 1. The structure of analysed websites focusing on "Business, Finance and Law" with user numbers

Source: own work

This provided an answer to research question Q1, which translates into attaining target T1.

The research used the quotations of the WIG[1], WIG20[2], mWIG40[3] and sWIG80[4] indexes of the Warsaw Stock Exchange. Such a choice resulted from the research format, i.e. searching for words without taking into account their context or their correlation with names of individual companies.

The following words were searched for in the analysed content: bear market, bull market, fall, rise, bear, bull, green, red, profit, loss (lose), recovery and crisis. They were selected ex-ante by the author. Their choice was influenced by the words' ability to reflect the sentiment and emotions on the capital market. They were confronted with the analyses conducted at 5:30 pm in order to determine the correctness of forecasting the direction of index change on a given day.

Then, using the sentiment analysis, the words were grouped into two classes (Table 1) that evoke negative (down) or positive (up) emotions.

Table 1. Division into positive and negative classes

Negative class (*Kn*)	Positive class (*Kp*)
bear market, fall, bear, red, loss, crisis. Polish word: bessa, spada, niedźwiedź, czerwony, strata, kryzys	bull market, rise, bull, green, profit, recovery. Polish word: hossa, rośnie, byk, zielony, zysk, ożywienie

Source: own work

All the words were declined and conjugated. Table 2 shows such forms for the Polish word 'strata' (loss).

Table 2. Forms of the 'strata' word

word: strata
declension forms: strata, straty, strat, stracie, stratom, stratę, stratą, stratami, stratach, strato

Source: own work

While answering research question Q2 and attaining target T2 it was analysed whether the occurrence of selected keywords at 5:30 pm corresponded to index changes on a given day. The results of analysis for selected 5 days and the WIG index change set against NOI'_j is presented in Table 3.

---

[1] The WIG index is the first Polish stock exchange index that has been calculated since 16 April 1991. The starting value of the WIG index was 1,000 points. Today WIG covers all companies listed on the WSE Main Market that meet the basic criteria for being included in indexes.

[2] The WIG20 index has been calculated since 16 April 1994 based on a portfolio of shares in 20 biggest and most liquid companies from the WSE Main Market.

[3] The mWIG40 index, calculated since 31 December 1997, has replaced the MIDWIG index and comprises 40 medium-cap companies listed on the WSE Main Market.

[4] The sWIG80, calculated since 31 December 1994, has replaced the WIRR index and comprises 80 small-cap companies listed on the WSE Main Market.

Table 3. The result of the analysis NOI'_j and its relation to the WIG index changes between 22 September 2015 and 28 September 2015.

Survey date	22 Sept 2015	23 Sept 2015	24 Sept 2015	25 Sept 2015	28 Sept 2015
IndexChangeClosing'$_{WIG,j}$ (pts)	−819.47	−516.31	−519.5	−50.87	−656.63
NOI'$_j$	Down	Up	Down	Down	Down
Value successfully/unsuccessfully reflected (pts)	819.47	−516.31	519.5	50.87	656.63
Cumulative value successfully/unsuccessfully reflected (pts)	819.47	303.16	822.66	873.53	1,530.16

Source: own work

The results of the whole analysis covering 280 days and the WIG, WIG20, mWIG40 and sWIG80 indexes are presented below.

Table 4. The results of the analysis NOI'_j covering 280 days and the WIG, WIG20, mWIG40 and sWIG80 indexes

Index name	WIG	WIG20	mWIG40	sWIG80
IndexChangeClosing'$_{WIG,j}$ successfully reflected (pts)	64,667.51	3,138.16	4,571.34	10,095.14
Total volatility *IndexChangeClosing'*$_{WIG,j}$ (pts)	95,400.13	4,655.52	7,128.31	16,569.59
Success ratio	68%	67%	64%	61%

Source: own work

The analysis suggests that the occurrence of keywords and the proposed division into classes forecast changes in the stock indexes in a better way than the toss of a coin. For each of the analysed indexes, the analysis showed more than 50% effectiveness in forecasting the direction of index change. This may confirm the suggestion that the selected positive and negative words and the proposed division into classes describe changes to stock indexes to an acceptable level.

To answer research question Q3 and attain target T3, press information that appeared before the start of trading was converted into a quantitative format, as presented in Table 5.

Table 5. Occurrence matrix in August 2016

String of words on day:	Frequency of occurrence, all grammatical forms of particular words														
	bear market	down (fall)	bear	overpriced	crisis	red	negative class value	bull market	up (rise)	bull	underpriced	recovery	green	positive class value	NOI
1 Aug 2016	1	120	2	0	20	0	143	1	49	4	0	5	5	64	D
2 Aug 2016	1	129	1	0	16	0	147	0	132	4	0	3	0	139	D
3 Aug 2016	0	189	0	0	22	1	212	1	217	1	0	0	0	219	U
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
29 Aug 2016	1	113	1	0	17	3	135	7	105	2	0	0	19	133	D
30 Aug 2016	7	144	1	0	13	1	166	10	229	0	0	2	0	241	U
31 Aug 2016	0	177	0	0	7	2	186	3	341	0	0	0	0	344	U

Source: own work

Then the result of class analysis was compared to the directions of stock index changes. The result of a 5-day analysis as compared to the WIG index changes is presented below.

Table 6. The result of class analysis and the directions of stock index changes for $IndexOpeningChange_{WIG,j}$ 22 September 2015 to 28 September 2015

Survey date	22 Sept 2015	23 Sept 2015	24 Sept 2015	25 Sept 2015	28 Sept 2015
$IndexOpeningChange_{WIG,j}$ (pts)	33.42	−37.02	60.31	284.65	−46.72
NOI_j	Down	Down	Down	Down	Down
Profit/loss resulting from forecast (pts)	−33.42	37.02	−60.31	−284.65	46.72
Cumulative profit/loss (pts)	−33.42	3.6	−56.71	−341.36	−294.64

Source: own work

The results of a full analysis for indexes WIG, WIG20, mWIG40 and sWIG80 for 9:00 am – trading starts, and for 5:00 pm – trading closes, are presented in Tables 7 and 8.

Table 7. The result of analysis NOI_j for 9:00 am, covering 280 days and indexes WIG, WIG20, mWIG40 and sWIG80

Index name	WIG	WIG20	mWIG40	sWIG80
$IndexChangeOpening_{w,j}$ correctly forecast (pts)	23,193.5	699.35	1,385.06	4,066.72
Total volatility $IndexChangeOpening_{w,j}$ analysed period (pts)	42,228.7	1,458.48	2,423.66	7,621.5
Forecast success ratio	55%	48%	57%	53%

Source: own work

Table 8. The result of analysis NOI_j for 5:00 pm, covering 280 days and indexes WIG, WIG20, mWIG40 and sWIG80

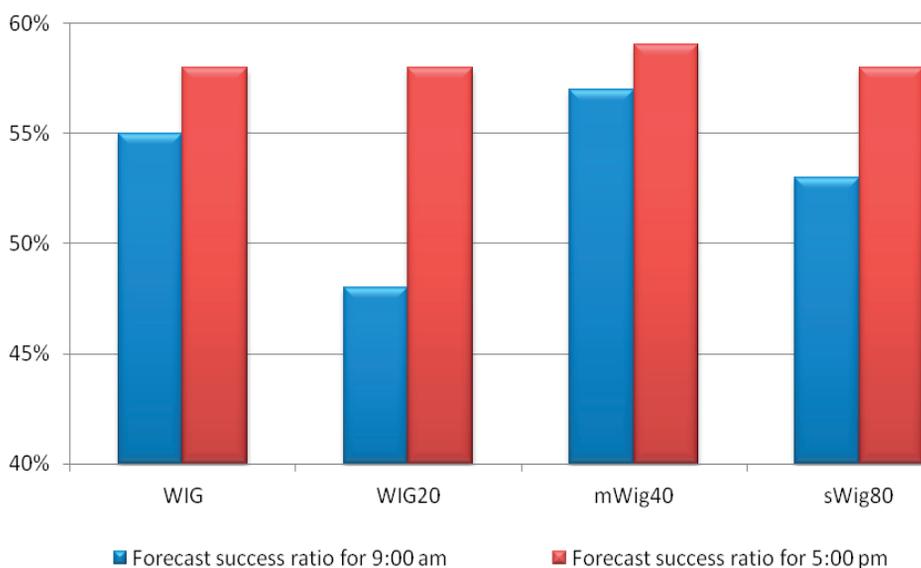| Index name	WIG	WIG20	mWIG40	sWIG80
$IndexChangeClosing_{w,j}$ correctly forecast (pts)	46,179.5	2,444.47	2,506.86	8,292.06
Total volatility $IndexChangeClosing_{w,j}$ analysed period (pts)	80,242.71	4,214.67	6,135.46	14,386.24
Forecast success ratio	58%	58%	59%	58%

Source: own work

Figure 2. The forecast success ratio for 9:00 am and 5:00 pm

Source: own work

Based on the analysis, it could be observed that the success ratio for the 9:00 am forecast is above 50% for the WIG, mWIG40 and sWIG80 indexes, and falls below that level, to 48%, only for the WIG20 index. The research of *NOI* and the changes in the stock index values for 9:00 am make it possible to draw a conclu-

sion that the Information Environment Sentiment before the opening of the trading does not significantly affect stock index movements when trading starts. In the case of the 5:00 pm analysis, all stock indexes achieved a success ratio significantly above 50%, higher than at 9:00 am for each of the indexes, as illustrated in the figure below.

This means that the Information Environment Sentiment before trading starts has a stronger impact on index changes at 5:00 pm comparing to 9:00 am. A conclusion can be drawn that investors making a buy/sell decision are more likely to incorporate available information at 5:00 pm rather than at 9:00 am, despite the fact that the information is already available on the websites before the start of trading.

5. Comments on the research results

The research results made it possible to attain the defined targets and answer the research questions. The primary research conclusion is the fact that there is a relation between online information and index changes on the Warsaw Stock Exchange. The attained efficiency of forecast at a level no lower than 58% makes it possible to acquire financial benefits on the capital market. Therefore further research under market conditions is required. If the efficiency of the forecast could be maintained at a similar level, the tool could be used as a basis for constructing an algorithm-based transaction system or for supporting decisions made by stock brokers. The research results suggest that the proposed solution could be used to assess the sentiment in the investors' online information environment as an alternative to the Investor Sentiment Index prepared by the Individual Investors Association.

The source of information of the designed tool does not have to be only financial-related websites, which contain mainly information provided by the Polish Press Agency (PPA), as well as comments and articles of analysts on the current economic and market situation. Information from social networking sites such as Facebook or Twitter is an alternative solution. Their users often regularly share information from their surroundings. Sometimes they witness a plane crash, a railroad accident, a terrorist attacks, earthquakes or other adverse events related directly or indirectly to companies listed on the Warsaw Stock Exchange. Such information is first disseminated among users of social networking sites, and only later is it sent to investors in the form of official stock exchange announcements, often after the collapse of the prices of the companies concerned.

Popularization of tools enabling the stock market investor to analyze large data sets in near real time will have a positive impact on reducing information barriers. As a result, there will be an increase in the rate of inclusion in prices of emerging information extracted from the analyzed data streams. Their selection, reading, understanding and interpreting by traditional methods, by an individual investor

is increasingly time-consuming. It is therefore possible to formulate a hypothesis that the implementation of analytical supportive solutions would improve the information effectiveness of the Warsaw Stock Exchange.

It should also be stated that the analytical tool ought to be further developed, with one of the fundamental issues being the identification of keywords and their division into classes. It is extremely difficult and involves detailed research among capital market participants as to the selection of words, their division into classes and setting the weightings levels. To this end, desk research and CAWI surveys should be carried out on stock investors. Determining the "absorption" time of online information from the investors' information environment also requires further research. As evidenced in the study, when trading starts at 9:00 am the information is not incorporated by investors to such a degree as when the trading closes at 5:00 pm. The time window probably features a point at which the forecast success ratio reaches its maximum level. Further analysis is suggested in order to find answers to the following research questions:

1. What keywords should be selected and how should they be divided into classes to maximise the efficiency of the forecast?
2. What is the optimum time within the trading day for forecasting the direction of stock index change?

The research will be another stage in expanding the author's knowledge and competence on the relationship between online information and index changes on the Warsaw Stock Exchange.

References

Das S., Chen M. (2001), *Yahoo! for Amazon: Extracting market sentiment from stock message boards*, [in:] Proceedings *of the Asia Pacific finance association annual conference*, vol. 35, Bangkok.

Dutta S. (2013), *Business Communications*, PHI Learning Private Limited, Delhi.

Hagenau M., Liebmann M., Neumann D. (2013), *Automated news reading: Stock price prediction based on financial news using context-capturing features*, "Decision Support Systems", vol. 55, pp. 685–697.

Hilbert M. (2012), *How much information is there in the "information society"?*, "Significance", vol. 9(4), pp. 8–12, doi:10.1111/j.1740–9713.2012.00584.x.

Ling R. (2012), *Taken for Grantedness: The Embedding of Mobile Communication into Society*, The MIT Press, Cambridge.

Loughran T., McDonald B. (2011), *When is a Liability not a Liability? Textual Analysis*, *Dictionaries, and 10-Ks*, "The Journal of Finance", vol. 66, no. 1, pp. 35–65.

Luhn H.P. (1958), *The automatic creation of literature abstracts*, "IBM Journal of Research and Development", vol. 2, pp. 159–165.

Lupiani-Ruiz E., García-Manotas I., Valencia-García R., García-Sánchez F., Castellanos-Nieves D., Fernández-Breis J.T. (2011), *Financial news semantic search engine*, "Expert Systems with Applications", vol. 38, pp. 15565–15572.

Mittermayer M.A. (2004), *Forecasting intraday stock price trends with text mining techniques*, [in:] *Proceedings of the 37th annual Hawaii international conference on system sciences*, Big Island.

Nassirtoussi A.K., Aghabozorgi S., Ying Wah T., Chek Ling Ngo D. (2015), Text *mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment*, "Expert Systems with Applications", vol. 42, pp. 306–324.

Nasukawa T., Yi J. (2003), *Sentiment analysis: Capturing favorability using natural language processing*, [in:] *Proceedings of the Conference on Knowledge Capture (K-CAP)*, Sanibel Island.

Nielsen F.Å. (2011), *A new ANEW: Evaluation of a word list for sentiment analysis in microblog*, [in:] M. Rowe et al. (eds.), *Proceedings of the ESWC2011 Workshop on „Making Sense of Microposts": Big things come in small packages 718 in CEUR Workshop Proceedings*, Heraklion.

Pang B., Lee L. (2008), *Opinion mining and sentiment analysis*, "Foundations and Trends in Information Retrieval", vol. 2, no. 1–2, pp. 1–135.

Peramunetilleke D., Wong R.K. (2002), *Currency exchange rate forecasting from news headlines*, "Australian Computer Science Communications", vol. 24, pp. 131–139.

Schumaker R.P., Chen H. (2009), *Textual analysis of stock market prediction using breaking financial news: The AZF in text system*, "ACM Transactions on Information Systems", vol. 27, pp. 1–19.

Tetlock P.C., Saar-Tsechansky M., Macskassy S. (2008), *More than words: Quantifying language to measure firms fundamentals*, "The Journal of Finance", vol. 63, pp. 1437–1467.

Tong R.M. (2001), *An operational system for detecting and tracking opinions in on-line discussion*, Working Notes of the SIGIR Workshop on Operational Text Classification, New Orleans.

Wirtualnemedia (2015), *Najpopularniejsze serwisy tematyczne w styczniu 2015*, http://www.wirtualnemedia.pl/artykul/najpopularniejsze-serwisy-tematyczne-w-styczniu–2015-roku# [accessed: 5.02.2017].

Badanie wpływu informacji sieciowych na zmiany indeksów na Giełdzie Papierów Wartościowych w Warszawie

Streszczenie: W artykule zaprezentowano wyniki badania nad wpływem informacji sieciowych pochodzących z serwisów internetowych o tematyce finansowej na zmiany indeksów zachodzące na Warszawskiej Giełdzie Papierów Wartościowych. Pierwsza część pracy ma charakter teoretyczny. Przybliżono w niej zagadnienie text miningu oraz analizy sentymentu. Przedstawiono ich zastosowanie w procesie analizy tekstu. W następnej części pracy omówiono charakterystykę prowadzonego badania. Dokonano wyboru polskich serwisów informacyjnych o tematyce finansowej, które mogą wpływać na reakcje inwestorów z Warszawskiej Giełdy Papierów Wartościowych. Przeprowadzono selekcję słów występujących w analizowanych serwisach oraz dokonano ich podziału na klasy. Następnie zaanalizowano zależności między zmianą indeksów GPW a częstością występowania poszczególnych słów w ramach klas. W ostatniej części pracy zaprezentowano wyniki badań, przeprowadzono dyskusję nad możliwościami ich wykorzystania oraz wskazano dalsze kierunki badań.

Słowa kluczowe: wiadomości, Giełda Papierów Wartościowych, text mining, analiza sentymentu

JEL: G12 G14