



Available for free online at <https://ojs.hh.se/>

Journal of Intelligence Studies in Business Vol 5, No 2 (2015) 36-47

Evaluation of e-Word-of-Mouth through Business Intelligence processes in banking domain

Lucie Šperková, Petr Škola, Tomáš Bruckner

Department of Information Technologies, Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic

lucie.sperkova@vse.cz

petr.skola@vse.cz

tomas.bruckner@vse.cz

Received September 1, accepted October 3 2015

ABSTRACT: Social networks and Internet discussions are valuable sources for a company's marketing research and public relations management. The Internet is full of public communication in an unstructured form and reflects recent movements of contributors' perception of the company, brand, products, competitors or whole market. As one of the approaches to achieve a better view we propose to design metrics which should be followed in order to get valuable insight where the company stands in terms of its customers. This paper focuses on obtaining an e-Word-of-Mouth in the banking sector using publicly available data. The main goal is to design metrics and dashboards evaluating customers' perception of a bank's services based on the analysis of public Facebook sites and web discussions related to several banks in the Czech Republic. We studied several approaches to unstructured data analysis. Thus we present complementary findings in classification of the unstructured data analysis presentation as a set of summarised metadata, top peaks of primary qualitative data and results of automated semantic analysis of the unstructured data. Based on the result we discuss the possible value of an unstructured data analysis and related systems. We find out that the value could be in the identification of opportunities and threats in the market by unexpected movements in public opinion of the Internet crowd, which we suggest to explore in future research. The benefit of this report is to describe the processing of data that can be obtained with emphasis on their content, their further enrichment, and their users.

KEYWORDS: Marketing, Business Intelligence, e-Word-of-Mouth, Elasticsearch, banking, unstructured data, Internet discussion, Facebook

1. Introduction

The phenomenon that people talk and recommend their favorite products and services to their friends and followers plays an important role in shaping their behavior (Goyette et al. 2010). Deeper understanding of these talks may be crucial in creating a successful marketing strategy in online communities. The most common method for monitoring these data is now online monitoring. But there are another data from different sources as internal and external databases, CRM, ERP, etc. in Data Warehouses of the companies which could be put into context with the data from online communities.

In marketing field these talks of customers are known as Customer Voice or Word-of-Mouth (WoM). The unstructured data gained from the Internet is also known as digital (Hu et al. 2006), electronic (eWoM) (Choi and Scott 2012) or online Word-of-Mouth (Wu and Zheng 2012). The most developed definition of the eWoM which captures present and future development of communication more complexly is stated by Bronner and de Hoog (2010) : “Any statement – positive, negative or neutral – made by potential, current or former stakeholders about a product, service, company or person, which is made available to a multitude of people, organisations or institutions, via a digitally networked platform.”

Potential of eWoM data can be used to obtain information to a broader audience as companies, professionals, and retrospectively users themselves. Any consumer in the world can connect to the Internet and read the opinions of others. The emergence of various social media like blogs, micro-blogs, social networks, forums, online reviews etc. is an important step for Customer Voice research. Users share there their personal experience with the companies, products and services and those are then followed by their transfer to other users. Simultaneously these sources create opportunity for companies to be visible and convince customers by communication on the quality of its services. Dellacorras et al. (2007) noted that the practice of reviewing products online significantly increases the potential for an empirical understanding of eWoM marketing. Breazeale (2009) states that digital platform is changing our understanding and the essence of the eWoM meaning. While the articulated evaluation disappears shortly after they were spoken, and it is very difficult to capture and analyze it, online statement lingers long after it was written and is not necessarily spontaneous. It is also immediate and accessible by others. Similar to classic WoM research shown that in Internet environment eWOM may have “higher credibility, empathy and relevance to customers than marketer-created sources of information on the Web” as stated Gruen et al. (2006, p. 449).

The importance of WoM in shaping consumers’ attitudes and buying decisions led many researchers to examine its effectiveness in stimulating demand

within various industries. There are researches including WoM influence on the buying decision and sales (e.g. Senecal and Nantel 2004; Tsang and Prendergast 2009; Chevalier and Mayzlin 2003), quality control (Ashton et al., 2014) or user and service experience (Hedegaard and Simonsen 2013; Pai et al. 2012). There are studies investigating eWoM in social networks (Wu and Zheng 2012) and estimating the social influence of individual nodes in social networks (Lü et al. 2011). Choi and Scott (2012) focus on the relationship between the use of social networking, user social capital, sharing knowledge and eWoM. The result shows that the intensity of use of the networks linked with confidence and identification, which has a positive impact on knowledge sharing.

Study conducted by Almassawi (2015) proved that “WoM has a positive influence on the youth’s decision-making process when choosing where to open a bank account”. This result can also lead to the importance of the customer segmentation according to their characteristics and preferences which they share on social networks and the data the banks has in their Data Warehouse. Banks can thus connect the characteristics from the social profile with the customer’s behaviour. Banking services are one of the industry where analysis of WoM can be crucial to stay competitive in the financial market. The potential is in identification what clients attracts, how the trends in banking look like, what is necessary to improve in services and also in what is necessary to help the clients and how to communicate with them in the space of social networks. Lack of confidence in the banking services might also be the result of an increase in perceived risk, which can reduce customers’ willingness to use banking services (Aurier and Siadou-Martin 2007). WoM can be a competitive advantage through banks can increase acquisition of prospects and retention of customers. The study of Shirsaver et al. (2012) found that the major determinant factors of positive WoM are corporate image, relationship marketing, perceived value, perceived risk, satisfaction, and loyalty. There are also studies which put in context WoM, service quality and customer satisfaction of the banking services (e.g. Yavas et al. 2004; Lymperopoulos and Chaniotakis 2008).

WoM affects individuals’ decisions and influences organizations’ operations. It has very important implications for a wide range of management activities, such as:

- building brand and reputation,
- increasing conversions, i.e. sales transactions,
- acquiring and retaining customers,
- product development,
- Quality Assurance.

Also business managers start to pay attention to social networks communication and new type of Business Intelligence is emerging (Chen 2010). In Business Process Management bringing together the

worlds of structured and unstructured data can add significant value to the enterprise. It can help to find the priority clients, problems relating to products and services, customer sentiment, find the next best step in business, identify activities of the competitors and customers, their reactions, etc.

Tremendous strides were made in recent years to automate the analysis of unstructured text data. The problem of semantic analyses is that their results should be quantifiable. Complexities in the analysis of unstructured textual data often results in only minimal use of the data (Ashton et al. 2014). So it is necessary to find a way how to generate outputs consumable to service providers. We are convinced that due to established culture, knowledge and technologies in companies the new methods has to adapt as much as possible to end users. According to Adamala and Cidrin (2011) the Business Intelligence solution must be built with end users in mind, as they need to use it.

2. Motivation of the research

Today Data Warehouses of banks contain mostly structured data as an asset they can easily measure. Business Intelligence (BI) is primarily directed to the presentation and analysis of numerical business data. Reporting systems, commonly based on dashboards, prepare quantitative data based on metrics in a report-oriented format that might include numbers, charts, or business graphics (Kemper et al. 2004). According to Kimball (2010) the metrics from the point of BI view are expressed on the basis of dimensional modelling as indicators and their characteristics, analytical dimension and their characteristics and the relationship between dimensions and indicators. COBIT 5 emphasizes the importance of business metrics. Metric is meant as degree, the extent to which company management is satisfied with the contribution of IT to meet business strategy.

Dashboards are applications that allow to organize pre-selected key performance indicators (metrics) in a clear and intuitive graphical form (Pour et al. 2012). At dashboard metrics can be viewed from many dimensions, for immediate use in decision-making processes in the organization. For business users dashboards bring the visibility and clarity of all monitored metrics and their instant overview of improving or deteriorating. Thus users can immediately assess the plan or reality and save their time.

Management of unstructured data determines how efficiently the company will deal with their customers in the future. The danger threatens from the ignorance of unstructured data can be sorted from dissatisfied customers, very loud customers, rapidly rising costs for customer service and their departure to breaching trust in the organization, the customer knows more than its employees. The new approach allows companies to consolidate unstructured data to central Data Warehouse is able to communicate consistently through all channels. The customer then

feels that company knows him when he communicates with his counterpart, whether it's agent or vendor, or attends a customer portal. Also customer service operations at the same time can reduce costs while maintaining customer satisfaction.

Integration of unstructured and structured data were discussed on presentation level (e.g. Becker et al 2002) where structured data are accompanied with relevant texts. The structured data selected as a results of metrics viewed from different dimensions and relevant documents are presented side-by-side. Another integration exists on the level of extracting metadata from collections of unstructured data (e.g. Keith et al 2005; Sukumaran and Sureka 2006). Identifiers of the content items are treated as facts that are subject to analysis, whereas metadata fields (e.g., author, date of creation, length, and addressed product) are used for classification purposes and thereby act as analysis dimensions. This allows associate individual documents with numerical facts directly, based on shared dimensions and to investigate document frequencies, e.g., the number of documents that cover a certain topic and are connected to certain segment of customers.

An integrated framework of Business Intelligence with the inclusion of unstructured data was constructed by (Baars and Kamper 2008), but they focus more on classic enterprise data and data from CRM. They do not include eWoM as a possible source of data to BI process. We are convinced, that eWoM is specific source of data which has to be handled in specific way. Our intention is also unique with its focus to banking domain, which has specific requirements to business. This paper follows results and expands article of (Šperková 2014) and (Šperková and Škola 2015), where the first content analysis of banking data were conducted.

Our purpose is automation of the process of gaining the data and their pre-processing for further analysis. Automation can reduce cost and time-consuming, manual and comprehensive analysis conducted by people like reading posts and search links in them. It is not able to capture the full transfer of expertise that customers write anywhere on the Internet. But at least in monitored publicly available sources can be analysed topics that interested users. Furthermore, these themes can automatically evaluate categories of sentiment and thereby obtain the distribution of subjects with positive or negative customer experience.

There are many studies conducted to mine the sentiment and opinion from the WoM and using the computer aided methods like Latent Semantic Analysis (Ashton 2014) or Machine Learning Classification (Pai et al. 2012). These methods are well known but are uneasily to implement in service practice. For this purpose, the powerful tool Elasticsearch seems to be adequate. There are only a few academic articles, which use Elasticsearch in their research. These articles are focused on library science (Johnson 2013) and full-text searching (Divya and Goyal 2013) or big log data (Bai 2013). Textual data analysis was the part of these

elaborated at the Department of Information Technologies at the University of Economics in Prague this year. These thesis uses unstructured data as the input and Elasticsearch as a tool for data analysis. Methodology used in those thesis are well-conceived and executed but lacks business context.

The nature of unstructured data in contrast with structured data usually presented in BI solutions is different and its meaningful presentation may differ from usual BI dashboards. We discuss the possibility of measurement and dashboard presentation relevant to the nature of the data and its business importance.

3. Objective and methodology

The main objective of this research is to create a periodic review of the data evaluating banks according to the context in which their users speak about them on the Internet. Our approach is built on the methods used in BI and knowledge from unstructured data processing in BI. The insight will be given based on metrics which have to be defined on the base of Facebook and Web comments. After processing of information from those comments, metrics are counted and visualized on dashboards. The results is an overview of the sentiment of the talks about the bank in specific period and its position in monitored metrics compared to other banks in the market. The research is conducted as a case study and proof of concept which will be followed by other studies and anchored in a methodology. Our approach is conducted according to established Business Intelligence process (Kimball 2010) and data mining, eventually text mining methodology, specifically according to CRISP-DM (Chapman 2000) as the main aim of this integration

is effective customer retention management. The lifecycle of the CRISP-DM contains 6 steps – business understanding, data understanding, data preparation, modelling, evaluation and deployment. Compliance with these procedures we outlined basic methodology of the research as follows:

1. Identification of the Web pages and social network sites where regular information from customers and users of banking services can be obtained – business and data understanding
2. Creating a system that will ensure downloading of the necessary data from the Internet and storing them in repository – data preparation
3. Processing and data analysis – data preparation and modelling
4. Design of metrics and characteristics, which evaluate the bank from the customers' point of view – modelling and evaluation
5. Design of the dashboard for the visualising the metrics and more detailed information – evaluation and deployment

Result of this paper is a dashboard which serves to further actions which should lead to better decision making and increasing performance. Figure 2 shows a general model of the decision making process from the unstructured data used at this research. The findings will provide important insights into the business impact of social media and user-generated content - an emerging problem in Business Intelligence research. Further this model can be easily integrated to the traditional, on structured data based, BI process.

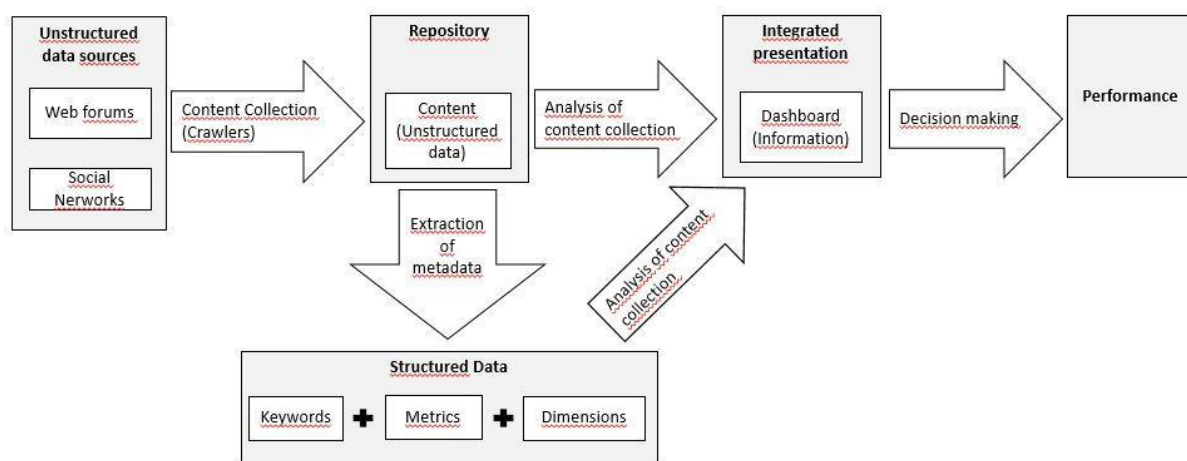


Figure 1: Model of the decision making process from unstructured data (authors)

3. Data collection and processing

From the marketing research point of view, East (2007) claims that it is not difficult to find the data on the Internet, but the problem may occur, if the data are only from one source/server. The eWoM may be affected and rather be negative or positive.

For this reason, we apply more than one data source. For the purposes of analysis and design of the metrics we chose comments that relate to banking occurring on the Czech website or Facebook profiles of Czech banks. Five Czech banks with the largest balance sheet total in 2012 and with the Facebook profile are shown in Table 1.

Table 1: Chosen banks with Facebook profile entering the analysis (authors)

<i>Name of the bank</i>	<i>Facebook profile</i>
Ceska sporitelna	ceskasporitelna
Komercni banka	komercni.banka
UniCredit Bank Czech Republic	UniCreditBankCZ
Raiffeisenbank	RaiffeisenbankCZ
GE Money Bank	gemoney.cz

Pages that a Web discussions are downloaded from must meet several requirements. Pages must relate to the topic of finances and banks to put assume a high proportion of the discussions that deal with banking services. Further discussions must have well-structured and tagged HTML code, so they can be easily identified in the whole HTML page script. From the Web sources we chose Czech financial forums <http://www.mesec.cz/> and <http://www.penize.cz/>.

4.1 Connectors

For downloading the data from the Internet forums we programmed a web crawlers for automatic browsing website content by using Java language and open source crawler4j library under the Apache Licence v2. In crawler4j we set up rules which domain to browse and optionally specified rules for browsing URLs that were interesting in their content. A list of text strings in the URL which should not be contained at pages was also defined for more efficient browsing. This crawler received information which parts of the site not to attend because they contain no user comments. Parts of the HTML code, containing identification of the contributor, text (comment), date of the comment and eventually the number of reviews of the comment by other users, were separated and prepared for further processing.

For acquisition of data from Facebook we used Java library RestFB which contains classes for working with Facebook objects. To login we used credentials (assigned App Id together with Access Token) for the application created below the private Facebook profile. The advantage of this log is access the data without the need to renew the validity of credentials. The objects of downloading from Facebook are posts on the wall of Czech banks and the data about the Facebook page which are downloaded from. Post can be represented by text, picture, link etc. Every post can contain comments from other users. The download these objects are accomplished by withdrawing feed objects first. For each object is determined whether contains a comment. If so, this comment is downloaded. Comments on Facebook are in two layers. For each comment, users can respond by sub-comments, these are downloaded as well. For each object type post is also necessary to determine the number of Likes - a positive evaluation of the object.

4.2 Repository

As a repository and analysis tool of gained data we used open source Elasticsearch software based on Apache Lucene library. Elasticsearch is a distributed scalable system for real-time search and analysis tool whose main function is the full-text search. It also supports structured search, geolocation and recording the relationships between data. In Elasticsearch, the data from all sources are collectively analysed. The data from both connectors are stored to Elasticsearch in JSON format. Every document contain unique identification under it is stored. This ID enables to start connectors over again each day because Elasticsearch saves one document under one ID. Downloaded data were enriched by other two Java programs, which connected sentiment analysis and evaluation of named entities contained in posts. Elasticsearch provides built-in support for analysis in the Czech language. Outputs from Elasticsearch were then processed and visualized in Kibana application. Plugin Head for simplification of indexes (data file) and application carrot2 for clustering documents were also used.

4.3 Sentiment Analysis

Sentiment analysis were conducted by open source OpenNLP library which is used for programming the various tasks of natural language processing like detection of sentences, tokenization, document categorization etc. Evaluation of sentiment contributions are made through OpenNLP Document Categorizer based on the principle of maximum entropy. For the training of the categorization model we used data from the University of West Bohemia as an output of sentiment analysis of data from the Czech Facebook sites and reviews from the Czechoslovak Film Database web using machine learning with a teacher.

5. WoM information extract

Before the design of metrics we explore which type of information could be extracted from the unstructured eWoM data. The successful BI initiatives, as shown in (Adamala and Cidrin 2011) share factors like orientation on choosing best opportunities (“low hanging fruit”) or alignment to specific needs of business sponsor. In our case the generic best opportunity could be found in fast, easy and simple understanding of movements in public opinion related to the company and its competitors. The business vision or specification of unstructured data analysis is difficult due the fact that the content of the data is not known in advance. Thus, the dashboard can be designed mostly by:

- summarised eWoM metadata,
- top peaks of eWom primary data,
- automated semantic analysis of eWoM data.

The metadata such as source, type or time of the contribution enables easy summarising and graphical representation. These data are easily integrable to current BI environments. The reason of these data in eWoM analysis is to understand time, typological and quantitative differences, and recent and past movements in eWoM data.

The metadata are source for identification of top peaks in primary data such as topics with the highest absolute or incremental rate of appearance or the persons with the high influence. These primary data should be shown to the dashboard user as a primary, non-summarised content, because it entails the semantics not easily evaluable by computer. For example, when rate of appearance of terms such as “availability”, “outage” or “failure” grows in conjunction with a competitor, it could be valuable information about technical conditions of competitor’s e-banking system. Also topics widely discussed about the company, e.g. social network campaigns started by unsatisfied customers, can be intercepted in its beginning.

The automated semantic analysis is represented mainly by sentiment analysis, ie. identification whether the contribution is neutral, positive or negative. The output to dashboard can be the quantity of the customer’s statements of different sentiment to measure the mood of the Internet public opinion or direct indication of sentiment of the top peak contributions. The example of a reason for semantic analysis could be an early cognition of negative or positive mood movements in the crowd after controversial marketing campaigns, thus is possible to avoid or intensify them.

6. Metrics design

The main purpose of metrics is to highlight the important facts that corporate resources or people need to be focused. Metrics summarize various aspects of the data in aggregate form and are comparable among the surveyed companies. From

downloaded and indexed data is necessary to draw metrics and other characteristics evaluating banks from the customers’ perspective. If the characteristics are of the quantitative type they are defined in proposed metrics in Table 2. Nominal characteristics are understood as dimension according them the metrics as measurable indicators can be calculated and sliced. Metrics along with dimension form the value for gaining the eWoM from the data. The highest value have dimensions created from textual analysis. Some results of metrics can be further used as dimensions to slice other metrics. For example one metric can be the *calculating sentiment* of different comments. Further this result can be used to slice metric *most active contributors* and show only those with negative sentiment. To better understand the content of posts and comments, the list of keywords has to be designed for better search of contributions according to user requirements. This is a domain knowledge of every enterprise which wants to use our procedure. This list can be always updated. Keywords are attributes for different dimensions. Considered dimensions in our case are:

- Time period (e.g. month, week, day, date)
- Source of the data (Facebook, Web forum)
- Type of the contributor (Facebook user, Bank, follower, user (cookie))
- Type of the page (individual Facebook page, individual forum page)
- Type of the contribution (comment, post)
- Name of the bank (keywords)
- Name of the product (keywords)
 - Specific
 - Generic
- Sentiment (positive, negative, neutral)
- Topic

Table 2: Definition of designed metrics (authors)

<i>Name of the metric</i>	<i>Definition</i>	<i>Calculation</i>	<i>Unit of measure</i>
Number of Likes	Indicates people who liked the page/post/comment, shows the popularity of the bank on Facebook	Summary of individual likes	Like
Number of posts	Shows the activity of the bank and its followers or other Facebook users, indicates how many objects of the type post are on the wall	Summary of individual posts	Post
Number of comments	Indicates how many objects of the type comment are on the wall/under the post	Summary of individual comments	Comment

The ratio of the number of comments that contain the name of the bank or the product to all comments	Evaluates how important it is to monitor the website and its discussion	Number of comments containing the specific topic / all comments	%
Frequency of the topics/keywords	Summarizes themes, a common signs of comments that occur most frequently	Summary of topics/keywords	Topic/Keyword
Incidence rate of topics together with keywords	Indicates topics occurring together with the keyword	Number of pairs of specific topic and the keyword/ number of occurrences of the keyword	%
Sentiment of the topic/contribution	Count the overall sentiment of the topic or contribution. Serves for example for comparison between banks.	(number of positive – number of negative contributions) / number of all contributions	Sentiment
Most active contributors	Users contributing the most – potential opinion makers	Summary of contributors	Contributor
Net Promoter Score	Evaluates measure of customer loyalty	% of loyal customers – % of disloyal customers	NPS
Reach	Summary of users who were reached by the post/comment	Summary of users who read the post/comment (number of impressions)	User/cookie

6. Dashboards design

Designed metrics need to be placed to the Dashboard. In our case the dashboard is realized in an application Kibana. Dashboard Overview shows defined metrics and contains a set of visualizations that correspond to the quantitative questions about

the stored data. Topic analysis dashboard shows topics or words that frequently occurred, or may be potentially interesting. It is designed to gain insight on the topics discussed in the context of the stored data. Dashboards are used for analysis of indexed data and are preparing for the final visualization.

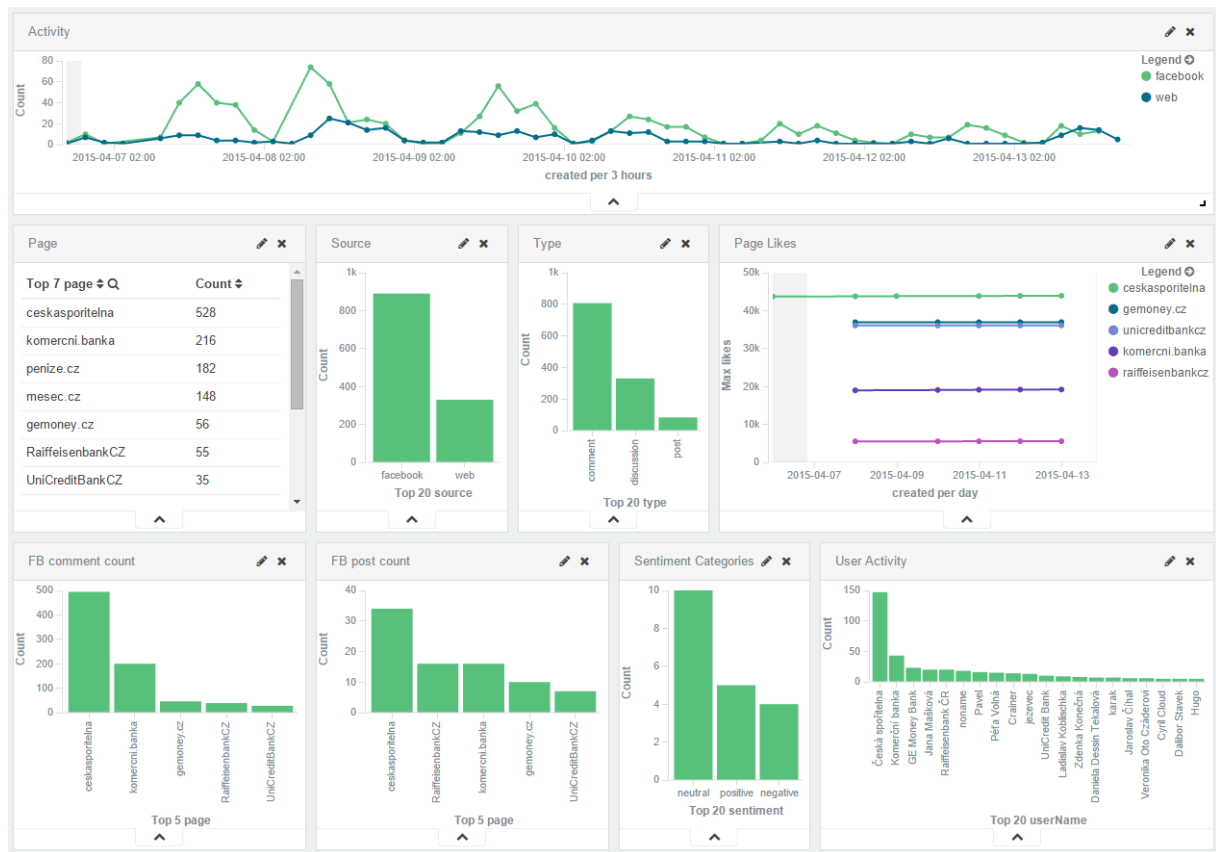


Figure 3: Preview of the Overview dashboard (authors)

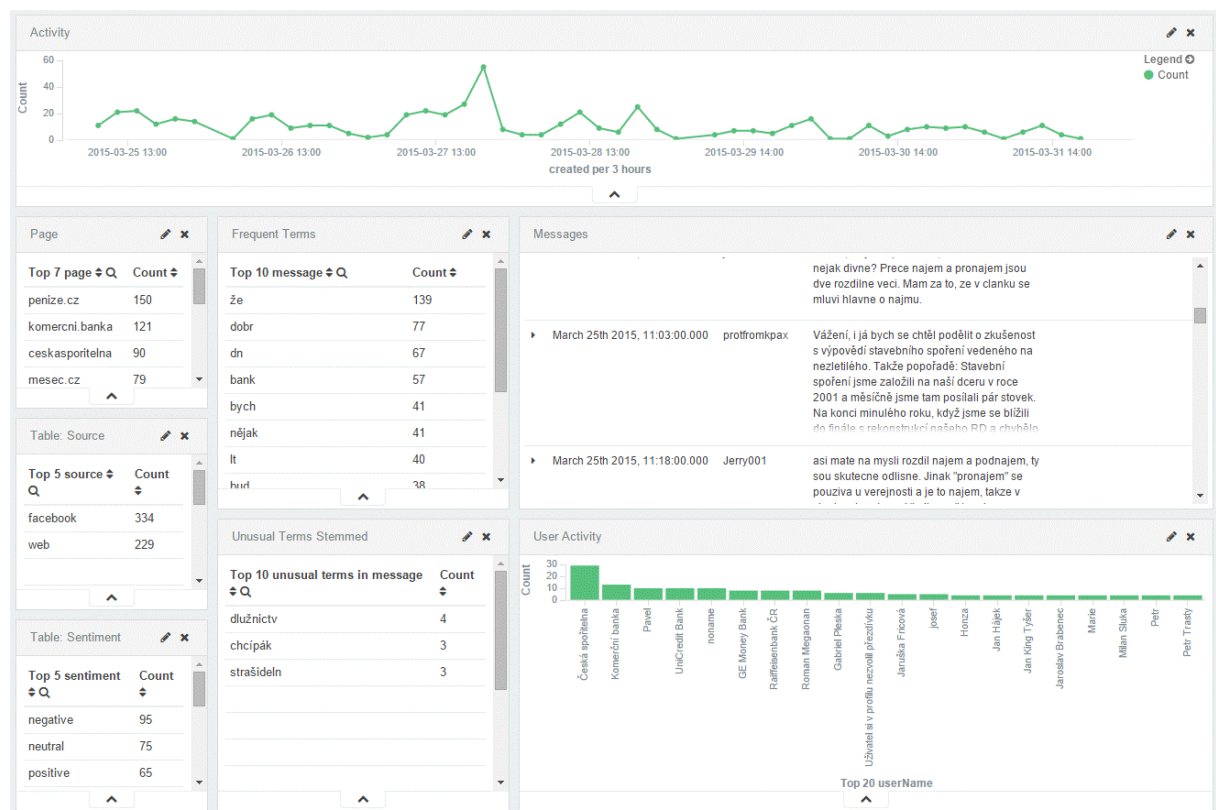


Figure 4: Preview of the Topic Analysis dashboard (authors)

Data can be viewed from different angles, search allows querying specific subsets of data. Data which contain the specific shapes of searched word or

phrase are then displayed. All objects defined in Table 5 and placed to dashboards also serve as filters that allow view data according to user interest. For

example, finding where there are many negative posts, which source caused a blip in the number of contributions etc. Another option is to enter a query into the search and thus, for example, determine whether the messages contained some of the key words or how often a name of the bank occurs.

Issues which were of interest of commenting can occur in several ways - with objects Frequent Terms and Top unusual terms and a frequency of posts in the course of time.

Table 3: Defined objects placed in dashboards (authors)

<i>Object</i>	<i>Description</i>	<i>Note</i>
Activity	Development of the number of posts in time	Filter allows to limit data, e.g. to period of high activity of users
Page	Names of the pages and number of documents	Filter by clicking on the name and change the sort by number of documents
Source	Number of documents from different sources	Filtering by sources
Type	Number of documents saved in single type of bank index	Filtering by the type of sources
Page Likes	Development of the number of Facebook likes in time	Popularity of sites can be compared between themselves
Comments count	Number of comments on individual pages	Filtering by contributors
Posts count	Number of posts on individual pages	Filtering by contributors
Sentiment	Summary of the sentiment evaluation	Sentiment is flow number from -1 to 1
User Activity	Most active users	Users with the highest number of comment or posts
Frequent Terms	Most frequent words contained in posts and comments	Identification of themes related to the contribution. It shows the word in the form after stemming rules and frequency of occurrence.
Top Unusual Terms	Terms that are statistically unusual	Terms which occur more frequently than they according to statistical model by other data should. It highlights the novelties in selected data

7. Further findings and implications of the study

The results of reporting design may serve as indicator of the marketing department for the evaluation of bank in relation to others in the market, as a feedback for new product introduction, overview of the competition or the discovery of the customer wishes. It indicates what bank is customer friendly and what bank and issues people talk about. Longer-term monitoring of metrics can therefore tell where to apply banking products. From a managerial perspective, our results suggest that firms should pay attention to textual content information when managing social media and, more importantly, focus

on the right measures. Therefore we also suggest closer cooperation of the people taking care of the social sites like Facebook and BI analysts. This approach could lead to higher customer satisfaction and growth of agility, profitability and orientation to the customers.

Though we consider the metrics and the dashboard design itself as a main result of our study, we are able to extract a typology of a possible information value and thus present a distinctive business value which could be requested from similar cases. We can also discuss the consideration of overall business value of unstructured data-based intelligence systems.

The following unstructured data value typology is made by observation of the data presented on the designed dashboards. Deployment of this typology to action in several situations is a base for future research. We suggest to perform a qualitative study by a sample of power users over the analysed data to authenticate the acceptance of the typology.

- Content / Quantity
- Primary qualitative data / Metadata
- Dynamics: Static (absolute) / Dynamic (change)
- Predictability: Expected / Surprising
- Object of business information value: Market / Competition / Customer / Product / Brand
- Business impact quality: Threat / Opportunity
- Source of information: Single / Nests / Distributed

We are able to find value even in the primary information itself, such as in the content of public contributions, or in the amount of the contributions or similar quantitative values. While the static data representing the absolute value are predictable, dynamic information representing changes - such as topics or words with the highest change rate of appearance - creates unexpected value. That leads to further exploration of reasons and origins of the information. Such origin sometimes lead to one source, sometimes to a/the competitor's campaign, sometimes to a single Internet personality with extended influence. Such an influencer could be a possible partner in public communication.

The overall business value of the unstructured data analysis is a sum of all of the expected business value described above. This makes similar systems very difficult to evaluate and to calculate a business case. A lot of value could be found in the area of unexpected, surprising information. It can create a big opportunity or prevent an extensive threat. Such value cannot be calculated in a simple business case, because it is impossible to set probability of a rise of such surprising information from the eWoM. Then the value and ROI of unstructured data intelligence systems could be considered similarly as in Business Continuity Management approach; as the avoiding the possible business impact of not having the information, eg. in our case, as possible business impact of ignorance of the Internet public opinion.

Conclusion

The purpose of the research was to design a comprehensive overview of customers' eWoM based on web forums and Facebook comments. After a study of the approaches to unstructured data analysis and WoM analysis, we discussed the nature of the unstructured data analysis and possibilities of its dashboard presentation. We defined quantitative metrics evaluating individual aspects of customers' perception of the bank, dimensions and the way they

can be displayed on the consolidated dashboards. We chose the Czech banking industry and Facebook pages and relevant websites with extensive discussions. The results were designed with respect to a possible future integration of the eWoM to Business Intelligence process and data structures in banks. The advantage of our approach is its extensibility. Connectors can be added for new sources of data; new metrics can be defined and incorporated to the dashboard. This approach can be also used besides banking in other enterprises.

The main outcome is the design of the metrics and the dashboards over the analysed public banking market data. The main findings are the way of the unstructured data analysis presentation as a set of summarised metadata, top peaks of primary qualitative data and results of automated semantic analysis of the unstructured data - especially the sentiment analysis, designed in the specific banking data dashboard. Furthermore, we discussed, generalised and classified the possible value of unstructured data analysis and related systems. We found out that the value could be in the identification of opportunities and threats within the market by unexpected movements in the public opinion in the Internet crowd, which we suggest to explore in future research.

In the case of positive results of the typology validation, the future research could contain automatic classification of the data to identify the type of business value of information presented on the dashboard and thus transfer more intelligence from humans to automated unstructured data processing.

8. Acknowledgements

This article was prepared with the financial support of the research project VSE IGS F4/18/2014 and with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.

References

- Adamala, S., Cidrin, L. 2011. Key Success Factors in Business Intelligence. *Journal of Intelligence Studies in Business*, Vol 1, No 1, pp 107-127.
- Almossawi, M.M., 2015. The Impact of Word of Mouth (WOM) on the Bank Selection Decision of the Youth: A Case of Bahrain. *International Journal of Business and Management*, Vol. 10, No. 4, pp 123-136.
- Ashton, T., Evangelopoulos, N., & Prybutok, V. R. 2014. Quantitative quality control from qualitative data: control charts with latent semantic analysis. *Quality & Quantity*, pp 1-19.
- Aurier, P., & Siadou-Martin, B. 2007. Perceived justice and consumption experience evaluations: A qualitative and experimental investigation. *International Journal of Service Industry Management*, Vol. 18, No. 5, pp 450-471.

- Baars, H., & Kemper, H.-G. 2008. Management Support with Structured and Unstructured Data—An Integrated Business Intelligence Framework. *Information Systems Management*, Vol 25, No. 2, pp 132–148.
- Bai, J. 2013. Feasibility analysis of big log data real time search based on Hbase and Elasticsearch. In: Natural Computation (ICNC), 2013 Ninth International Conference on. IEEE, 2013. pp 1166–1170
- Becker, J., Knackstedt, R., & Serries, T. 2002. Informationsportale für das Management – Integration von Data-Warehouse- und Content-Management-Systemen. In Vom Data Warehouse zum Corporate Knowledge Center—Proceedings der Data Warehousing, Physica-Verlag HD, pp 241–261.
- Bronner, F. & de Hoog, R. 2010. Consumer-generated versus marketer-generated websites in consumer decision making. *International Journal of Market Research*, Vol. 52, No. 2, pp 231–248.
- Dellarocas, C., Zhang, X. & Awad, N. 2007. Exploring the value of online product reviews in forecasting sales: the case of motion pictures. *Journal of Interactive Marketing*, Vol 21, No. 4, pp 23–45.
- Divya, M. S. & Goyal, S. K. 2013. An advanced and quick search technique to handle voluminous data.
- East, R. 2007. Researching word of mouth. *Australasian Marketing Journal (AMJ)*, Vol. 15, No. 1, pp 23–26.
- Goyette, I., Ricard, L., Bergeron, J., & Marticotte, F. (2010). e-WOM Scale: word-of-mouth measurement scale for e-services context. *Canadian Journal of Administrative Sciences*, 27(1), pp 5–23.
- Gruen, T. W., Osmonbekov, T., & Czaplewski, A. J. 2006. eWOM: The impact of customer-to-customer online know-how exchange on customer value and loyalty. *Journal of Business research*, Vol 59, No. 4, pp 449–456.
- Hedegaard, S., & Simonsen, J. G. 2013. Extracting usability and user experience information from online user reviews. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, pp 2089–2098.
- Hu, N., Pavlou, P. A., & Zhang, J. 2006. Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of Online word-of-mouth communication. Proceedings of the 7th ACM conference on Electronic commerce, ACM, June, pp 324–330.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. 2000. CRISP-DM 1.0 Step-by-step data mining guide.
- Chen, H. 2010. Business and market intelligence 2.0. *IEEE Intelligent Systems*, Vol 25, No. 1, pp 68–71.
- Chevalier, J.A. & Mayzlin, D. 2006. The effect of word of mouth on sales: online book reviews. *Journal of Marketing Research*, Vol 43, No. 3, pp 345–354.
- Choi, J. & Scott, J., 2013. Electronic Word of Mouth and Knowledge Sharing on Social Network Sites: A Social Capital Perspective. *Journal of theoretical and applied electronic commerce research*, Vol 8, No. 1, pp 11–12.
- Jašek, P. 2014. Analyzing User Activity Based on RFM Models Complemented with Website Visits and Social Network Interactions. IDIMT-2014 (Networking Societies – Cooperation and Conflict), Poděbrady, 10.09.2014 – 12.09.2014, Linz : Trauner, ISBN 978-3-99033-340-2, pp 181–190.
- Johnson, T., 2013. Indexing linked bibliographic data with JSON-LD, BibJSON and Elasticsearch. *Code4lib Journal*, Vol. 19, pp. 1–11
- Keith, S., Kaser, O., & Lemire, D. 2005. Analyzing Large Collections of Electronic Text Using OLAP”, 29th Conf. Atlantic Provinces Council on the Sciences (APICS 2005), Wolville (Canada), October.
- Kemper, H. G., Mehanna, W., & Unger, C. (2004) “Business Intelligence—Grundlagen und praktische Anwendungen”, Wiesbaden: Vieweg.
- Kimball, & R.Ross, M. 2010. The Kimball Group reader : relentlessly practical tools for data warehousing and business intelligence. Indianapolis: Wiley Publishing, ISBN 978-0-470-56310-6.
- Lü, L., Zhang, Y. C., Yeung, C. H., & Zhou, T. 2011. Leaders in social networks, the delicious case. *PloS one*, Vol. 6, No. 6, e21202.
- Lymperopoulos, C., & Chaniotakis, L. 2008. Price satisfaction and personal efficiency as antecedents of overall satisfaction from customer credit products and positive word of mouth. *Journal of Financial Services Marketing*, Vol. 13, No. 1, pp 63–71.
- Pai, M. Y., Chu, H. C., Wang, S. C. & Chen, Y. M., 2013. Electronic word of mouth analysis for service experience. *Expert Systems with Applications*, Vol. 40 No. 6, pp 1993–2006
- Pour, J., Maryška, M., & Novotný, O. 2012. Business Intelligence v podnikové praxi. Professional Publishing, ISBN 978-80-7431-065-2.
- Senecal, S. & Nantel, J. 2004. The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, Vol. 80, pp 159–169.
- Shirsaver, H., Gilaninia, S., & Almani, A. 2012. A study of factors influencing word of mouth in the Iranian banking industry. *Middle-East Journal of Scientific Research*, Vol. 11, No. 4, pp 45–460.
- Sukumaran, S., & Sureka, A. 2006. Integrating Structured and Unstructured Data Using Text Tagging and Annotation. *Business Intelligence Journal*, Vol 11, No. 2, pp 8–17.
- Šperková, L. 2014. Analýza nestrukturovaných dat z bankovních stránek na sociální síti Facebook. *Acta Informatica Pragensia*, Vol 3, No. 2-spec., pp 154–167.
- Šperková, L., Škola, P. 2015. Design of Metrics for e-Word-of-Mouth Evaluation From Unstructured

- Data for Banking Sector. In: 16th European Conference on Knowledge Management. Udine, 03.09.2015 – 04.09.2015. Udine : Published by Academic Conferences and Publishing International Limited Reading, pp 717–725. ISBN 978-1-910810-46-0.
- Tsang, A. & Prendergast, G. 2009. Is a ‘star’ worth a thousand words? *European Journal of Marketing*, Vol. 43, pp 1269–1280.
- Wu, W., & Zheng, R. 2012. The impact of word-of-mouth on book sales: review, blog or tweet?. In Proceedings of the 14th Annual International Conference on Electronic Commerce. ACM, August, pp 74-75.
- Yavas, U., Benkenstein, M., & Stuhldreier, U. 2004. Relationships between service quality and behavioral outcomes: A study of private bank customers in Germany. *International Journal of Bank Marketing*, Vol. 22, No. 2, pp 144-157.