Web Reputation Index for XU030 Quote Companies

Bilal Cankir
Department of Business, Kirklareli University, Kirklareli, Turkey
Email: bilalcankir@klu.edu.tr

M. Lutfi Arslan and Sadi Evren Seker
Department of Business, Istanbul Medeniyet University, Istanbul, Turkey
Email: lutfi.arslan@medeniyet.edu.tr, academic@sadievrenseker.com

***Abstract*—Aim of this study is researching the web based reputation index of most prestigious companies quoted in the Istanbul stock market, first time. The web reputation is gathered through the 12 different parameters, collected from Google, Facebook, Twitter, Bing, Alexa, etc. All the parameters are normalized by min-max method to achieve a common value range. The parameters, and methodology is explained together with the web based reputation index of 30 companies.**

***Index Terms*—reputation, web-o-metric, data mining, business intelligence**

I. Introduction

Aim of this study is building a web based reputation index for the XU30 quote companies. The XU30 companies are the companies with highest stock market values among 417 companies[1]. All the companies quoted to the stock market should have their web pages and publish the details of their company financial statements and annual, by law.

By this study, it is first time researched the correlation between the stock market values of quotes and the web based reputation in Turkey.

The research can be considered as a new trend on electronic reputation (e-reputation) and as a part of reputation index based on web pages, feedbacks and some social networks [1].

Electronic reputation has an increasing trend especially on electronic market places (e-commerce) and electronic business of companies (e-trade) is supposed to be followed by the companies [2].

The same index value is calculated for each of the 30 companies in the Istanbul stock market. The final results vary from 0.0 to 0.4 and the companies with highest web reputation are mostly from banking sector while the lowest rates are from the mining sector. The relation of sector and web reputation is obvious, since the banking sector has a web interface for the customers, while mining sector mostly deals with business-to-business (B2B) operations and both the number of users are relatively low and the number of web interactions are also relatively low.

As it is demonstrated on Fig. 1, the web indicators of a web page of all the companies are gathered from the Internet resources. In this study, 12 different parameters are gathered from 8 different sources. After the collection of values by indexer in Fig. 1, the parameters are normalized to get combined into a single index. The normalized index values are considered as the web reputation index of the XU030 quotes.

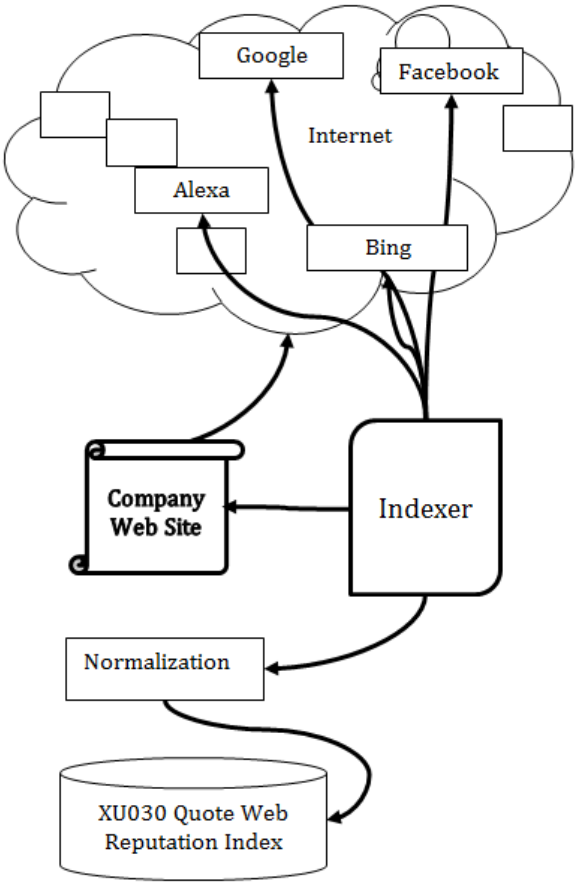

Figure 1. Data flow diagram of the study

Manuscript received January 20, 2014; revised July 2, 2014.

[1] Number of companies in BIST (Istanbul Stock Market) is 417 by the date Oct, 2013.

II. Background

Employing specific web mining techniques, researchers try to set relations between different variables [3], and measure the impact and effectiveness of web sites including academic ones. Quantitative aspects of web data and information have created separate sub-disciplines like informetrics, cybermetrics, and webometrics. Informetrics is "the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists" [4], and cybermetrics is "the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the whole Internet drawing on bibliometric and informetric approaches" [5].

Webometrics, a term first coined by Almind and Ingwersen, is a measurement of the effectiveness of web sites [6]. According to Thelwall, it is "the study of web-based content with primarily quantitative methods for social science research goals using techniques that are not specific to one field of study" [7]. First example of this measurement is the "Web Impact Factor" (WIF) developed by Ingwersen and defined as "the number of web pages in a web site receiving links from other web sites, divided by the number of web pages published in the site that are accessible to the crawler" [8].

Similar techniques, which have been developed throughout the years, all showed that there is a significant correlation between financial statement of companies and the reputation of web sites [9]-[12]. Especially, a study by Márkmol and Kuhnen showed that external in-links correlated general ratings of the companies [13]. For business companies, same kind of relation was also found. There is a significant correlation between the number of in-links to the web site of a company and its business performance [14], [15].

The proposition behind web-o-metrics is that web visibility and impact of a university is highly correlated with its reputation. Those universities with good reputation tend to have more visible web sites, high traffic, more links etc.

There are five set of tools of web-o-metric research: link analysis, web citation analysis, search engine evaluation, descriptive studies of the web, and the analysis of Web 2.0 phenomena. While link analysis measures the hyperlinks between web pages, web citation analysis counts how often journal articles are cited. Search engines are used to evaluate the extent of the coverage of the web and the accuracy of the reported results. Descriptive studies include various survey methods like the average web page size, average number and type of meta-tags used, the average use of technologies like Java and JavaScript, the number of users, pages and web servers. Last but not least tool is Web 2.0 applications [16]-[21].

As the aim of our paper, we use tools of webometric research like Google page rank, number of visitors, number of pages linking back to the web page or the number of likes on Facebook, in order to create a reputation index. Our intention is to be as simple and usable as possible. We can summarize our method as follows.

III. Data Mining

The data mining phase is built over 12 different parameters collected from different web metric sites like alexa, bing, google, facebook or twitter. This chapter explains the details of the parameters.

Has a Facebook Page? We have checked whether the companies have a Facebook page or not. Fortunately all of the companies have a Facebook page except one so we have removed this metric from our calculations.

Facebook like Count: Facebook is the leading social network with highest number of members around the world. We have collected the number of like counts for each of the companies. The maximum like count is for Turkcell and the count is 2.747.255. The minimum value is 0 for the company without the Facebook page. The average value for the Facebook like count is 273.693 and the reason of high standard deviation can be related to the Facebook campaigns of some companies. For example the highest Facebook like count company is a well-known telecom company with the Facebook campaigns.

Value of the Site: Some of the independent organizations offers a free agent to calculate the expected value of the web site via the web indicators like Alexa ranking or Google page rank. Most of them are built on the number of visitors and expected click from the visitors to make a valuing.

BING Backlinks: The BING back links are collected from the search engine of Microsoft, Bing. The maximum number of Bing backlinks is 3540 for Akbank and the average number of back links is 137.

Google Backlinks: Google backlink count is the number of page sites indexed by the Google crawler. This number is under the effect of two facts. First, the number of pages held on the web site is limited. For example if a web site has only 1 page, the maximum possibility for the Google backlinks is 1. Second, even the web page can hold multiple pages, Google can crawl only a part of the web pages. The maximum number of back links is 3.313.000 for TurkTelekom, while the average is 307.817 for all 30 companies.

Daily Unique Visitors is the average number of visitors per day. The daily visitor number can vary from date to date and we have collected the up to date values during the research time. The maximum visitor is 637.285 for Garanti Bank and the average for 30 companies is 62.656.

Alexa Ranking is another indicator published by an Amazon owned web site alexa.com. The lesser number means the web page has a higher ranking and the minimum ranking for the web site is 24 in Turkey and highest ranking is 65.836 among the whole Turkish web sites on the Internet. Another parameter is the Alexa global ranking, similar to the Turkish ranking. The lowest global ranking is 1.442 and the average is 570.013 among all the web sites on the Internet.

Time on Site is a web indicator to measure the time spending of the users with a time interval of their entrance and exit. The higher time means a higher

reputation for the web site and the maximum value of time spent on the web page is about 8 minutes and average is about 4 minutes. These time intervals are also daily, which means the time on site indicator is an average day based time spending on web page for each of the user.

Facebook Shares is another indicator that is the count of the shares of the web site of the company. The value is fetched from Facebook and the higher number of shares is considered as a positive indicator for the company web site. Unfortunately the numbers available for public access on Facebook is limited with last month. So the number of shares are only limited with last 30 days. The average share count is 211 and the maximum count is 1969 for Turkcell.

Tweets parameter is the count of tweets mentioning about the web site of the company. Again, similar to the Facebook shares, the publicly available tweets are limited. Maximum number of tweets is 276 for Halkbank and the average number is 22.

Google Trend is the publicly available trend calculator built on the Google search data. Trends values can be both queried as a time series or as the latest value of the trend [21]. We have also added the google trends values for each of the companies in BIST30 as their brand values. The trend values of the brands vary from 19 to 100 where the 100 is the maximum available and 0 is the minimum possibility of the google trends.

IV. Normalization

In the normalization phase, the collected web indicator values are normalized via min-max normalization [19].

$$N_{Min\,Max}(x) = \frac{x - Min}{Max - Min} \quad (1)$$

The normalized value is calculated by the subtraction of the minimum value of the series from the sample and dividing the subtraction to the distance between minimum and maximum values of the series.

The reason of normalization is getting comparable values for each of the indicators. For example, some of the web parameters have values up to millions while some are only limited to 100. For this reason we need a common scale for all of the parameters and we have implemented the min-max normalization for each of the parameters where the result is between 0 and 1.

Another problem in combining multiple parameters into a single metric is the effect of parameters as positive or negative direction. For example the Alexa ranking of a web site can be considered as a negative directed effect on the combination, since the better reputation comes from smaller rankings. As a solution we have calculated the inverse of these indicators by multiplying with -1, which means a subtraction in the final decision in fact.

So the total score is calculated with below formula.

$$WRI = \frac{\sum_0^C N_x - \sum_C^K N_x}{C} \quad (2)$$

The Web Reputation Index (WRI) is calculated with the summation of negative indicators subtracted from the summation of positive indicators divided by the count of positive indicators "*C*". The "*K*" symbol in above formula stands for the total number of indicators which is the summation of positive and negative indicator counts.

Because the summation of positive indicators is always higher than the summation of negative indicators the equation of WRI always gets a positive real number between 0 and 1.

V. Results

This section holds the details of the normalized index values. The complete list of universities with the index values are placed into the appendix of the paper.

Properties of the data set are given in Table I.

TABLE I. Properties of the Index Values

Mean (μ)	0.202
Maximum	0.392
Minimum	0.065
Standard Deviation (σ)	0.085
Total Number of Companies	30

The distribution of the university reputation index is given as a separate figure.

In Fig. 2, the x-axis holds a unique number for each of the university, while the y-axis demonstrates the normalized web reputation value.

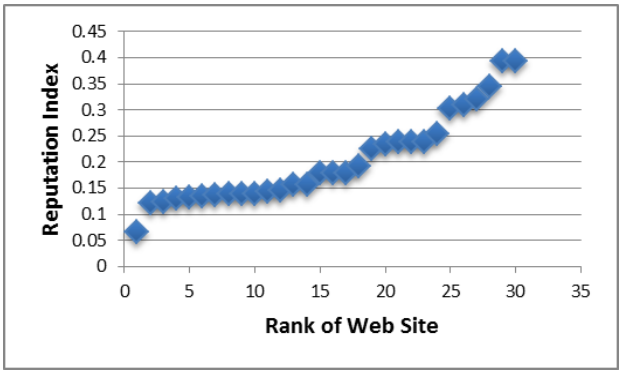

Figure 2. Statistical distribution of normalized web reputation index

VI. Conclusion

In this study, the most prestigious companies quoted to the Istanbul Stock Market are studied from the web reputation perspective. By the law, all the companies quoted to the stock market should have a web page for at least publishing their financial status to the shareholders. The results show that some of the companies have higher impact of reputation on their web pages while some of the companies are relatively less active and their web pages are less repute. We believe this study will guide most of the reputation studies especially related to the web-o-metrics.

APPENDIX NORMALIZED WEB REPUTATION INDEX FOR XU030

Company	WRI
KOZA MADENCİLİK	0.065857677
KARDEMİR	0.121338154
KOZA ALTIN	0.123361817
DOĞAN HOLDİNG	0.13080305
ASELSAN	0.133048538
SABANCI HOLDİNG	0.134191004
ERDEMİR	0.137845855
MİGROS	0.138208073
IHLAS HOLDİNG	0.138510134
ENKA İNŞAAT	0.139221673
TAV HAVA MEYDANLARI	0.143908579
PETKİM	0.14551056
SİŞECAM	0.15587265
EMLAK KONUT	0.156934331
KOÇ HOLDİNG	0.178281699
TOFAŞ	0.178351108
ARÇELİK	0.179148381
TÜPRAŞ	0.192145463
BİM	0.224913252
PEGASUS	0.234415601
BANK ASYA	0.237515509
THY	0.238876722
YAPI KREDİ	0.239586477
VAKIFBANK	0.254795463
TÜRK TELEKOM	0.303086627
AKBANK	0.308522205
HALKBANK	0.321051624
TURKCELL	0.345335151
GARANTİ	0.39235097
İŞBANK	0.392439757

REFERENCES

[1] G. E. Bolton, E. Katok, and A. Ockenfels, "How effective are electronic reputation mechanisms? An experimental investigation," *Management Science*, vol. 50, no. 11, pp. 1587-1602, 2004.

[2] G. Zacharia, A. Moukas, and P. Maes, "Collaborative reputation mechanisms for electronic marketplaces," *Decision Support Systems*, vol. 29, no. 4, pp. 371-388, 2000.

[3] S. E. Seker, C. Mert, K. Al-Naami, U. Ayan, and N. Ozalp, "Ensemble classification over stock market time series and economy news," in *Proc. IEEE International Conference on Intelligence and Security Informatics*, 2013, pp. 272-273.

[4] J. Tague-Sutcliffe, "An introduction to informetrics," *Information Processing & Management*, vol. 28, no. 1, pp. 1-3, 1992.

[5] L. Björneborn, "Small-world link structures across an academic Web space: A library and information science approach," Ph.D. dissertation, Royal School of Library and Information Science, Copenhagen, Denmark, 2004.

[6] T. C. Almind and P. Ingwersen, "Informetric analyses on the world wide web: Methodological approaches to 'webometrics'," *Journal of Documentation*, vol. 53, no. 4, pp. 404-426, 1997.

[7] M. Thelwall, "Introduction to webometrics: Quantitative web research for the social sciences," *Synthesis Lectures on Information Concepts, Retrieval, and Services*, vol. 1, no. 1, pp. 1-116, 2009.

[8] P. Ingwersen, "The calculation of web impact factors," *Journal of Documentation*, vol. 54, no. 2, pp. 236-243, 1998.

[9] K. Mcnally, M. P. O'mahony, M. Coyle, P. Briggs, and B. Smyth, "A case study of collaboration and reputation in social web search," *ACM Transactions on Intelligent Systems and Technology*, vol. 3, no. 1, pp. 1-29, 2011.

[10] I. O. Williamson, J. E. King Jr., D. Lepak, and A. Sarma, "Firm reputation, recruitment web sites, and attracting applicants," *Human Resource Management*, vol. 49, no. 4, pp. 669-687, 2010.

[11] I. Caraganciu, "Market theories analysis for the web market and incorporation of reputation as a competitive factor for web companies," *Studies in Business and Economics*, vol. 8, no. 1, pp. 29-35, 2013.

[12] I. Constantiou, N. Hoebel, and R. V. Zicari, "Reputation, framing strategies and user's choice of content on the web: An empirical study," *Concurrency and Computation: Practice and Experience*, vol. 22, no. 7, pp. 872-889, 2010.

[13] F. G. Mármol and M. Q. Kuhnen, "Reputation-Based web service orchestration in cloud computing: A survey," *Concurrency and Computation: Practice and Experience*, 2013.

[14] L. Vaughan and G. Wu, "Links to commercial websites as a source of business information," *Scientometrics*, vol. 60, no. 3, pp. 487-496, 2004.

[15] L. Vaughan, "New measurements for search engine evaluation proposed and tested," *Information Processing & Management*, vol. 40, no. 4, pp. 677-691, 2004.

[16] M. Thelwall, "Bibliometrics to webometrics," *Journal of Information Science*, vol. 34, no. 4, pp. 605-621, 2008.

[17] S. E. Seker and K. Al-Naami, "Sentimental analysis on turkish blogs via ensemble classifier," in *Proc. International Conference on Data Mining*, 2013, pp. 10-16.

[18] S. E. Seker, "Sentimental versus impact of blogs," in *Proc. International Conference on Internet Computing and Big Data, ICOMP*, 2013, pp. 127-133.

[19] S. E. Seker, K. Al-Naami, and L. Khan, "Author attribution on streaming data," in *Proc. IEEE 14th International Conference on Information Reuse and Integration*, 2013, pp. 497-503.

[20] M. L. Arslan and S. E. Seker, "Web based reputation index of Turkish universities," *International Journal of E-Education E-Business E-Management and E-Learning*, vol. 4, no. 3, pp. 197-203, 2014.

[21] S. E. Seker and B. Diri, "TimeML and Turkish temporal logic," in *Proc. IC-AI*, 2010, pp. 881-887.

Bilal Çankır was born in Istanbul in 1985. He applied to a MA program (International Quality Management) at Marmara University. He finished this school at 2010. He's studied PhD at Istanbul University (Management) since 2011 autumn. He is currently working at Kirklareli University Faculty of Economics and Administrative Sciences in Business Administration area since August 2010.

Mehmet Lütfi Arslan was born in Vezirkopru in 1972. He completed his undergraduate business degree in Marmara University of Istanbul. He obtained his MA and PhD in Social Sciences Institute of Marmara University. During his postgraduate studies, he worked in private sector. He is currently Assistant Professor of Management and Organization in newly founded Istanbul Medeniyet University and studies on management, organizations, entrepreneurship, and human resources management.

Şadi Evren Şeker was born in Istanbul in 1979. He has completed his BSc., MSc. and PhD. Degrees in computer science major. He also holds an M.A. degree in Science Technology and Society. His main research areas are Business Intelligence and Data Mining. During his post-doc study, he has joined data mining research projects in UTDallas. He is currently Asst. Prof. in Istanbul Medeniyet University, Department of Business. He is an IEEE member and senior member of IEDRC.