

# Big Data Analytics Study «Third Wave»

Olga Kolesnichenko / 1st  
Security Analysis Bulletin  
Moscow, Russia  
[oykolesnichenko@list.ru](mailto:oykolesnichenko@list.ru)

Victor Dashonok / 3rd  
EMC Academic Alliance Russia & CIS  
St. Petersburg, Russia  
[victor.dashonok@emc.com](mailto:victor.dashonok@emc.com)

Lev Mazelis / 5th  
Department of Mathematics and Modeling  
Vladivostok State University of Economics and Service  
Vladivostok, Russia  
[lev.mazelis@vvsu.ru](mailto:lev.mazelis@vvsu.ru)

Gennady Smorodin / 2nd  
EMC Academic Alliance Russia & CIS  
St. Petersburg, Russia  
[gennady.smorodin@emc.com](mailto:gennady.smorodin@emc.com)

Oleg Zhurenkov / 4th  
Department of Mathematics and  
Engineering computer science in Economics  
Altai Academy of Economics and Law  
Barnaul, Russia  
[zhur@pie-aael.ru](mailto:zhur@pie-aael.ru)

Dariya Yakovleva / 6th  
Department of Information Technology and Systems  
Vladivostok State University of Economics and Service  
Vladivostok, Russia  
[darya.yakovleva15@vvsu.ru](mailto:darya.yakovleva15@vvsu.ru)

**Abstract**— In this study EMC Academic Alliance for the first time has applied the new approach to organize the Big Data Analytics study. The Multicenter Study «Third Wave» was presented at the EMC Academic Forum Russia & CIS 2015, at the International Scientific Congress «Globalistics-2015» and at the Russia National Supercomputing Forum 2015. Was implemented the variant of morphological analysis of Text Big Data from the Internet. The Study «Third Wave» is provided by Bill Schmarzo's 6 Steps Approach. Countries are set into positive or negative image's groups, and also into two groups with indication to technological order. The study can be developed to elaborate a new index for countries' rating. Big Data Analytics Study «Third Wave» relates to the new sphere for investigation – determination of Global management metrics.

**Keywords** — *Big Data; Big Data Analytics; HPDA; Social Network; Index of Economic Freedom; Networked Readiness Index; Fragile States Index.*

## I. ACADEMIC AD-HOC TEAM

EMC Academic Alliance Russia & CIS has launched in 2015 the Big Data Analytics Multicenter Study «Third Wave» with participating of three universities: Lomonosov Moscow State University, Altai Academy of Economics and Law and Vladivostok State University of Economics and Service. The study has been named after Alvin Toffler's «Third Wave» concept about information era [1]. In this study EMC Academic Alliance for the first time has applied the new approach to organize Big Data Analytics study – Academic ad-hoc Team, the volunteer academic research and education cooperation in open mode.

Students have had the practical experience implementing the laboratory works, they included the results into their Master's degree theses. One of the important aspects of such cooperation was the preparation of articles to scientific journals

and to the international conferences [2, 3]. The Study «Third Wave» was presented at the EMC Academic Forum Russia & CIS 2015, at the International Scientific Congress «Globalistics-2015» dedicated to 70-year anniversary of the United Nations and at the Russia National Supercomputing Forum 2015.

## II. HPDA IS AVAILABLE FOR EVERY STUDENT

### A. High Performance Data Analytics

EMC Academic Alliance serves as the open platform for all stakeholders among high schools of Russia, CIS region and beyond. Alliance is the part of the wider global network of academic partnerships of EMC Corporation, bringing together universities of the USA, India and other countries. The main goal of EMC Academic Alliance is to shape the future labor market and creation a high skilled workforce oriented to a new trends and technologies [4, 5]. Having lack of accessibility to supercomputer technology for students, greatly increases the need for High Performance Big Data Analytics (HPDA) in all spheres of the economy [6]. In order to meet the need for HPDA, participated in the Study «Third Wave» universities used the multi-level hybrid schemes of Data Mining. For Data Mining were used: Google and Yandex API; R-STUDIO; GitHub; Microsoft SQL 2012 Enterprise Edition. Google and Yandex are considered as non-classical hybrid supercomputers with API access «as-a-Service».

The experience of organizing and providing Multicenter Study «Third Wave» by EMC Academic Alliance clearly showed that every student and teacher have the opportunity to participate in HPDA. Lack of such skills is critical for universities' graduates and in general it forms negative impact on economy innovation development. But lack of high computer capacity in a university cannot justify the fact that

students will not acquire practical skills in Big Data Analytics during education period. The cooperation in the common framework of professional academic community such as EMC Academic Alliance allows every student to HPDA practice.

### B. Text Big Data Analytics as priority of HPDA

In the Multicenter Study «Third Wave» was implemented the variant of morphological analysis – count of specific keywords from Web pages and Social Networks depending on given conditions. Large arrays of unstructured text data in the Internet belong to the categories of Big Data and are suitable for rapid monitoring of the situation. Was shown that unstructured arrays of keywords reflect the real global society processes in the Internet. The accumulation's dynamics of keywords in the Internet reflects the situation in the countries.

### III. 6 STEPS APPROACH OF BILL SCHMARZO

It is important to carry research using a right approach for gaining the results. EMC Chief Technology Officer Bill Schmarzo has developed the 6 Steps Approach to the successful implementation of Big Data Analytics projects [7, 8, 9]. During implementation the Multicenter Study «Third Wave» was used Bill Schmarzo's 6 Steps Approach, that helped to gain the main findings.

#### Step 1. Identify Key Business Initiative

Was chosen Case-study of Big Data Analytics: the image of the countries in the global Internet. Was developed the hypothesis of assessment how countries' image and economic indicators are reflected in the Internet.

#### Step 2. Identify Strategic Nouns (entities)

Were selected keywords for Data Mining: terrorism, terrorist, occupation, narcotic, violation, democracy, development; drip irrigation, mobile phone, solar panel, nuclear power plant, oil. Based on dominant keywords were identified positive (Democracy, Development) and negative (Terrorist, Terrorism, Narcotic) image's groups of countries. And also were identified two groups of countries by economic trends 2008/2015: «Oil/Oil» and «Oil/IT».

#### Step 3. Brainstorm Strategic Noun Questions

It is needed to answer to three types of questions: descriptive, predictive and prescriptive.

*Descriptive Analytics* – assessment of the image of countries in the Internet for past 15 years. Countries were set into positive or negative image's groups.

*Predictive Analytics* – description of positive or negative forecast for countries. Countries were set into two groups: with indication to change of technological order 2008/2015 «Oil/IT» after 2008 year's world financial crisis or with indication to absence of change of technological order 2008/2015 «Oil/Oil». Countries in the group with indication to change of technological order have a better forecast (see part IV below).

*Prescriptive Analytics* – understanding what affects to positive or negative forecast for countries. For example,

Turkey, Tunisia and Libya are in positive image's group, but they have low indication to development of solar energy in comparison with other countries in the positive group, that worsens the forecast for these three countries. The prescriptive point is to develop solar energy and IT more.

#### Step 4. By Analysis

- Rows by keywords (morphological analysis).
- Rows by years from 2000 to 2015.
- Rows by countries' names: Afghanistan, Azerbaijan, Georgia, Israel, Iraq, Iran, Yemen, Kyrgyzstan, China, Libya, Pakistan, Palestine, Syria, Tunisia, Turkey, Ukraine, Uzbekistan.

The selected rows of Big Data (by keywords, countries and years) allowed to understanding the clear picture related to elaborated hypothesis and also allowed to find the statistically significant correlations with some metadata [3, 5], that showed the whole picture for understanding of the relationship among the spread of mobile phones within population and the escalation of social unrest during «Arab Spring».

#### Step 5. Score Technique

Were taken three global indexes as a reference points:

IEF – Index of Economic Freedom (The Heritage Foundation) [10];

NRI – Networked Readiness Index (World Economic Forum) [11];

FSI – Fragile States Index (Fund for Peace) [12].

The task of developing a new index during the first phase of the study was not intended. The statistically significant difference was found only for Index of Economic Freedom between positive and negative groups of countries (Fig. 1).

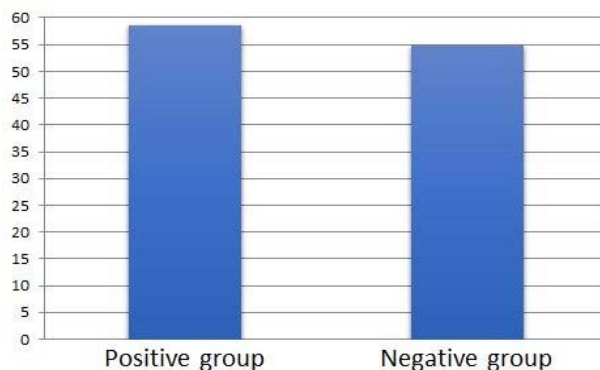


Fig. 1. The average of Index of Economic Freedom in positive and negative groups of countries, F-test  $p < 0,05$ .

*Step 6. Close The Loop*

Recommendations Worksheet included providing to the teams that were engaged into Data Mining, the detailed Datafication Codebook for implementation of research.

According to the results the study can be developed to investigate the social unrest prediction [3, 5]. And the study can be developed to elaborate a new index for countries' rating based on Text Big Data Analytics (Figures 2-5).

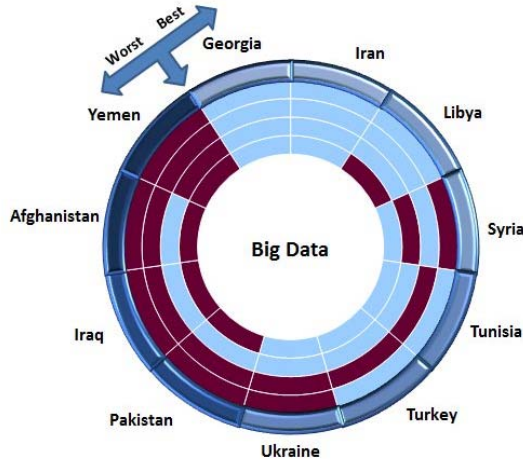


Fig. 2. The ranking of the countries by criteria identified in the Study «Third Wave». External first circle – colors from better forecast (light bars) to worse forecast (dark bars) for the countries. Second circle – positive (light bars) or negative (dark bars) groups of countries based on dominant keywords. Third circle – change of technological order to «IT» in 2015 (light bars) or absence of change of technological order «Oil» (dark bars). Fourth circle – growth of keywords «solar panels» from 2008 to 2015 (light bars) or absence of growth (dark bars). Internal fifth circle – growth of keywords «drip irrigation» from 2008 to 2015 (light bars) or absence of growth (dark bars).

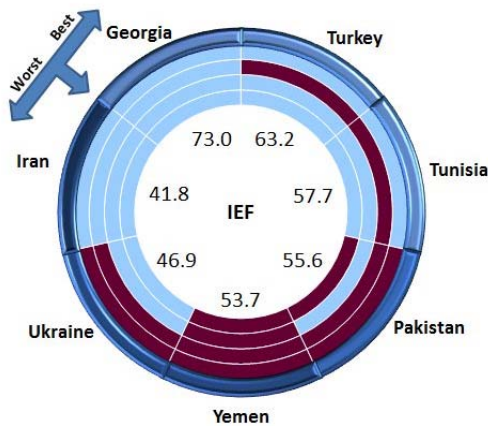


Fig. 3. Ranking of the countries by Index of Economic Freedom with criteria identified in the study «Third Wave», from better forecast (light bars) to worse forecast (dark bars). Circles and bars' color accord to criteria identified for every country in the study «Third Wave» (see Figure 1).

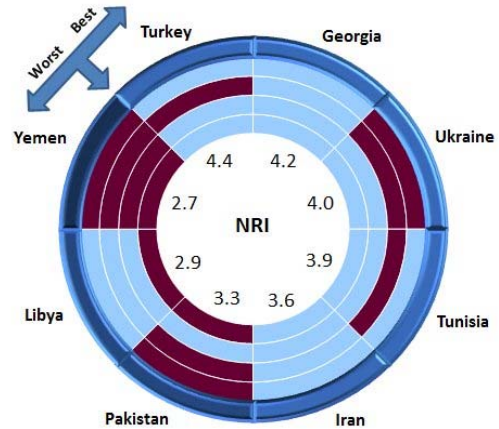


Fig. 4. Ranking of the countries by Networked Readiness Index with criteria identified in the study «Third Wave», from better forecast (light bars) to worse forecast (dark bars). Circles and bars' color accord to criteria identified for every country in the study «Third Wave» (see Figure 1).

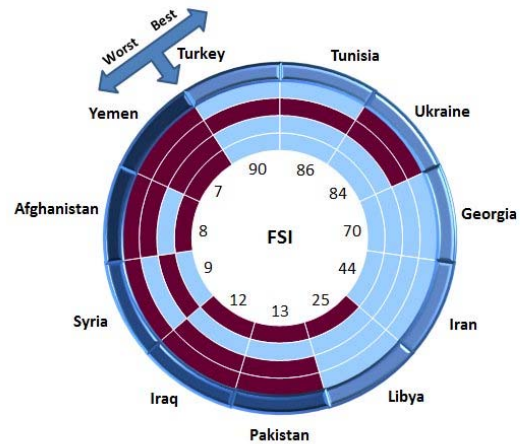


Fig. 5. Ranking of the countries by Fragile States Index with criteria identified in the study «Third Wave», from better forecast (light bars) to worse forecast (dark bars). Circles and bars' color accord to criteria identified for every country in the study «Third Wave» (see Figure 1).

Study showed that unstructured text Big Data – arrays of keywords in the Internet – reflect the real economic processes, in particular in energy sector. Dataficated keywords «solar panels» have positive correlation with statistical characteristic «GDP PPP /purchasing power parity» (metadata from IEF) and also with statistical characteristic «Electricity production, kWh/capita» (metadata from NRI). It means that solar energy has become a real sector of economy, it is included into main trends of states' total power generation and into states' revenues count.

#### IV. POSITIVE AND NEGATIVE FORECAST

As was mentioned above in this study, based on dominant keywords, were determined positive and negative image's groups of countries, and due to indication to change of technological order 2008/2015 «Oil/IT» or absence of change «Oil/Oil» was discussed the forecast for countries. It is not the comprehensive method for forecast using Text Big Data Analytics, but first attempt to build some model for forecast. (Figures 6, 7). The most demonstrative is the interconnection between dataficated dominant keywords from Data Mining.

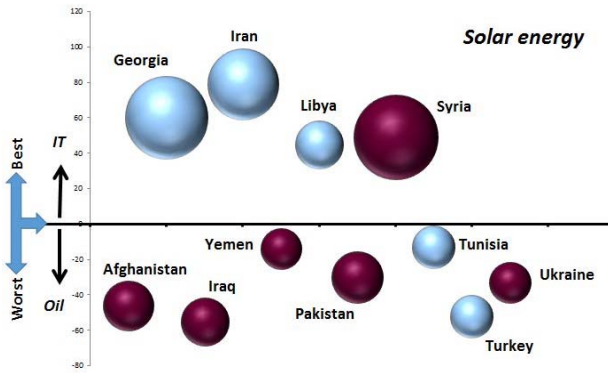


Fig. 6. Interconnection between dataficated dominant keywords from Data Mining. The size of the circles is number of keywords «solar panel» in millions. The axis Y is the number of keywords «mobile phone» and «oil» in a millions. Light circles belong to the positive group of countries (dominant keywords «democracy», «development»), dark circles belong to the negative group of countries (dominant keywords «terrorist», «terrorism», «narcotic»).

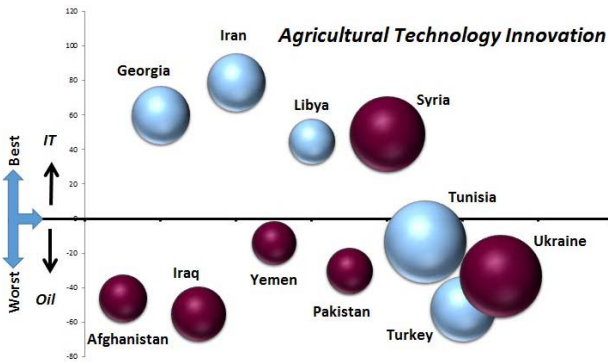


Fig. 7. Interconnection between dataficated dominant keywords from Data Mining. The size of the circles is number of keywords «drip irrigation» in millions. The axis Y is the number of keywords «mobile phone» and «oil» in a millions. Light circles belong to the positive group of countries (dominant keywords «democracy», «development»), dark circles belong to the negative group of countries (dominant keywords «terrorist», «terrorism», «narcotic»).

Were considered two important trends for future development and economic innovation: solar energy and agricultural technology innovation. Both apply to IT development. Countries belonging to the positive group (indicated by light circles) are characterized in the most cases by more number of keywords «solar panels» and «drip irrigation», and also in the most cases they belong to group with indication to change of technological order 2008/2015 «Oil/IT». The countries from negative group (indicated by dark circles) in the most cases are characterized by less number of keywords «solar panels» and «drip irrigation», and they belong in the most cases to group with absence of change of technological order 2008/2015 «Oil/Oil».

Using interconnection between dataficated dominant keywords can be shown which country has more or less capacity for innovation and advanced technology development, or which country has more or less terrorism problem, and also which country is more or less prepared to the new energy market and which country has more or less agriculture (food) efficiency.

During next stage of Multicenter Study «Third Wave» will be elaborated more keywords using Text Big Data Analytics for more wide model.

#### V. GLOBAL MANAGEMENT METRICS

In the chain of Big Data Analytics «Monitoring – Insights – Optimization» [8] before start to make the changes it should be estimated the current situation by score / metrics. During Globalization's process have appeared the definitions «Global Governance» or «Global management metrics». In terms of Big Data it concerns ranking of the countries by different indexes, which will be built on Big Data Analytics, including text's HPDA (morphological analysis and more complicated semantical analysis). Since the beginning of computer modelling for estimation of global processes were used different statistical metrics [13, 14]. Today the most discussed issue is the need to seek a new, more relevant metrics than just statistical GDP-related indicators [15]. It can be indicators of sustainable development, energy efficiency, renewable energy, innovation, nature conservation, human resources, Millennium Development Goals and more. Big Data Analytics as opposed to statistical methods allows to flexible and wide analysis. Also Big Data Analytics allows to estimate countries in accordance with Global system. There are three features which characterize the Global system: Non-equilibrium; Openness with information exchange; Nonlinearity [16]. Only Big Data Analytics can take account of these three features.

#### VI. CONCLUSION

In this article is discussed only part of all study results. But all main findings from first phase of Big Data Analytics Multicenter Study «Third Wave» should be mentioned.

1) Based on Text Big Data Analytics (Web pages and Social Networks) countries can be set into groups: positive or

negative in terms of image; and with indication to change of technological order or with absence of it.

2) Was determined the post-crisis (after world financial crisis in 2008 year) change of technological order, which is characterized by the transition to the Information Technology and solar energy, and is associated with the countries' development and positive image of countries.

3) Was found the spatio-temporal structure (surface diagram of correlation field) as reflection of the political phenomenon «Arab Spring» in the Internet that has correlations with mobile phones distribution within population (it isn't considered in this article, see more in [3, 5]).

## VII. NEXT STEP. NEW GLOBAL CHALLENGES

Next step to elaborate that kind of open, multicenter study under EMC Academic Alliance coordination should be international cooperation based on search for a common values, uniting different nations factors and a new, more relevant metrics for ranking of the countries. India is the global laboratory for a new models of development and a new ideas for global market due to it's big population, democracy and flexibility of the economy [17]. That is why India is the priority for a new large-scale projects of Big Data Analytics. It is of interest to investigate Internet as open system of linguistics culture implementing Text Big Data Analytics.

## REFERENCES

- [1] E. Toffler, *The Third Wave*, Moscow, Russia: AST, 2004.
- [2] O. Kolesnichenko, G. Smorodin, V. Dashonok, B. Schmarzo, I. Ilyin, O. Zhurenkov, L. Mazelis, D. Yakovleva, "The topicality and experience of Big Data Analytics Multicenter Study provided by EMC Academic Alliance", in *Proc. International educational conference EMC Academic Forum 2015*, pp 5-17, Taganrog, Russia, 2015.
- [3] O. Kolesnichenko, G. Smorodin, I. Ilyin, O. Zhurenkov, L. Mazelis, D. Yakovleva, V. Dashonok, "Big Data Analytics. New opportunity to understand the global processes", in *Proc. IV International Scientific Congress Globalistics-2015*, dedicated to the 70th anniversary of the United Nations, Moscow, Russia, 2015.
- [4] G. Smorodin, "Workforce in 2020", in *Proc. EMC Academic Forum Russia & CIS*, pp 3-5, Moscow, Russia, 2014.
- [5] G. Smorodin, O. Kolesnichenko, "Big Data as the Big Game Changer, in *Proc. 9th IEEE International Conference on Application of Information and Communication Technologies*, pp 40-43, Rostov-on-Don, Russia, 2015.
- [6] D. Dietrich, B. Heller, B. Yang, *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*, USA: Wiley, 2015.
- [7] B. Schmarzo. (2015, April-June). Thinking Like a Data Scientist, Parts I-III, EMC Blog In Focus. Available: [https://infocus.emc.com/william\\_schmarzo/thinking-like-a-data-scientist-part-i/](https://infocus.emc.com/william_schmarzo/thinking-like-a-data-scientist-part-i/)
- [8] B. Schmarzo. (2014, January). Big Ideas Big Data Business Model Maturity Index), EMC Blog In Focus. Available: [https://infocus.emc.com/william\\_schmarzo/big-ideas-big-data-business-model-maturity-index/](https://infocus.emc.com/william_schmarzo/big-ideas-big-data-business-model-maturity-index/)
- [9] B. Schmarzo, *Big Data: Understanding How Data Powers Big Business*, USA: Wiley, 2013.
- [10] Index of Economic Freedom. The Heritage Foundation. Available: <http://www.heritage.org/index>.
- [11] Networked Readiness Index. "Global Information Technology Report", World Economic Forum, 2015. Available: <http://reports.weforum.org/globalinformation-technology-report-2015>.
- [12] Fragile States Index. Fund for Peace. Available: <http://global.fundforpeace.org>.
- [13] J. Forrester, *World Dynamics*, Moscow, Russia: AST, 2003.
- [14] S.P. Kapitsa, "Essay on the Theory of human population growth. Demographic revolution and information society", public lectures "Russia in the Global context", Moscow, Russia, 2008.
- [15] S.N. Bobylev, "Energy efficiency and climate: modernization approach", Lomonosov Moscow State University, Moscow, 2010.
- [16] I.V. Ilyin, O.G. Leonova, A.S. Rozanov, *Theory and practice of political globalistics*, Moscow, Russia: Publishing house of the Lomonosov Moscow State University, 2013.
- [17] O. Kolesnichenko, A. Rozanov. (2014, May). "Report: India as Global laboratory for Reverse Innovation", 1st Multidisciplinary Summit of the Global and International Research Associations, Lomonosov Moscow State University, Security Analysis Bulletin. Available: <http://www.securityanalysisbulletin.com/analysis/6/18/india-as-global-laboratory-for-reverse-innovation.htm>.