Trends and Patterns of Public Interest to Big Data in International Community

Tendencias y pautas de interés público para el big data en la comunidad internacional

Artur I. PETROV
ORCID: http://orcid.org/0000-0003-2634-0567
ID-Scopus: 57191265004
artlgpetrov@yandex.ru
Institute of Transport, Tyumen Industrial University, Tyumen, Russian Federation

Daria A. PETROVA
Institute of Natural Sciences and Mathematics, Ural Federal University named after the first President of Russia B.N. Yeltsin, Ekaterinburg, Russian Federation

Victor I. KOLESOV
Construction Institute, Tyumen Industrial University, Tyumen, Russian Federation

ABSTRACT

The issues of analysis of the dynamics of changes in public interest to the topic of Big Data in the world, the US and Russia are considered in the article. A research tool is a regression analysis of Google search analytics on the monthly activity of Internet users in the search for information on this topic. It is shown that, despite the relatively lower level of requests for Big Data in Russia, the growth rate of interest in this topic in Russia is higher than in the US. Phase analysis showed that the process under study is unstable.

Keywords: Analytical data; information technology; regression models; social interest.

RESUMEN

Las cuestiones de análisis de la dinámica de los cambios de interés público en el tema de Big Data en el mundo, los Estados Unidos y Rusia, se consideran en el artículo. Una herramienta de investigación es un análisis de regresión de Google Search analytics sobre la actividad mensual de los usuarios de Internet en la búsqueda de información sobre este tema. Se muestra que, a pesar del nivel relativamente más bajo de solicitudes de Big Data en Rusia, la tasa de crecimiento del interés en este tema en Rusia es mayor que en los Estados Unidos. El análisis de fase mostró que el proceso en estudio es inestable.

Palabras clave: Datos analíticos; tecnología de la información; modelos de regresión; interés social.

INTRODUCTION

The term "Big Data" was introduced by John R. Mashey, Chief Scientist, SGI in 1998 and was originally used to identify the socio-economic phenomenon associated with the emergence of technological capabilities for the qualitative analysis of huge data sets. It was fully accepted by the scientific community on September 4, 2008. It was then that a special issue of the authoritative journal Nature was published, entirely devoted to the future of technologies based on the processing of huge amounts of data (Campbell, 2008).

Nowadays, under the Big Data, we are meaning the set of approaches, instruments and methods for processing data of huge volumes and significant diversity (Agneeswaran, 2012; Avdeeva, 2016; Bhadani and Jothimani, 2016; Boyd and Crawford, 2012; Chernyak, 2011; Cukier, 2010; Jin *et al.*, 2015; Li and Lu, 2014; Manyika *et al.*, 2011; Meyer-Shenberger, 2014; Miller and Mork, 2013).

Big Data technologies are already the basis in technologies for analyzing information requests of all kinds (Chernyak, 2011; Meyer-Shenberger, 2014). A striking example – Internet search engines analyzing the statistics of visitors to various sites, their specificity and search for proposals in the field of interests (Agneeswaran, 2012; Cukier, 2010).

On the basis of the collection and processing of massive amounts of information, a variety of tasks are being solved, including the formation of appropriate political elections for voters (D. Trump during the US election campaign in 2016 (Von Grassegger and Krogerus, 2016)). In this connection, the question of tendencies (trends and patterns) of the development of Big Data technologies in Russia in comparison with the technologically advanced countries of the world is quite interesting.

In the modern world, the dynamics of technological change is extremely high. Annually there are innovative developments, which have all the chances for promotion due to venture funds (Agneeswaran, 2012; Avdeeva, 2016; Chernyak, 2011; Meyer-Shenberger, 2014). Not all of them in the end are successfully implemented; many ideas turn out to be foneys (Bhadani and Jothimani, 2016; Boyd and Crawford, 2012). In this regard, there was a need to track and then classify the stages of the life cycle "Idea – its study – investment – development – rush support – practical sobering – getting to the fork "development of techniques for practical use" or "oblivion". This is what the research and consulting company Gartner (Cukier, 2010; Gartner, 2015), specializing in information technology markets, does.

In the annual Maturity Surveys of IT-technologies prepared by Gartner specialists since 1995, Big Data was at the stage of a technological trigger for about 5 years (2008 – 2012), in 2013, they were on Peak of Inflated Expectation and since mid-2014 have entered the phase of getting rid of illusions (Trough of Disillusionment), when the shortcomings of technology are revealed, and the loss of novelty does not contribute to enthusiastic publications, and the community notes disappointment with the new technology.

For example, in an analytical review of 2015, Gartner no longer represents the Big Data technology on the Hype Cycle chart (Figure 1) (Gartner, 2015) even at the *Trough of Disillusionment* stage. According to the theory of technology development (Hype Cycle chart), then Big Data can wait for either a phase of oblivion, or the stages of *Slope of Enlightenment* and *Plateau of Productivity* (maturity of technology). The question arises, how fair is the disappointment (as of summer 2015) of Gartner's experts in Big Data technologies? Does this general disappointment in Big Data correspond to the situation in the Russian IT sector? The hypothesis that is proposed to be checked by query analytics in Google Trends can be formulated as following: "With the overall relatively lower level of development of Big Data technologies in Russia at the moment, interest in this technology in our country is higher than in the US and in the world".

Figure 1: Big Data topic fallout of the number of emerging technologies
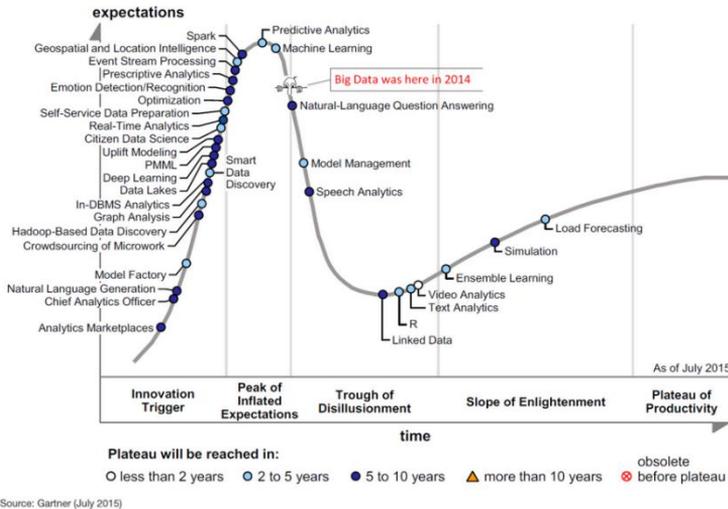
METHODOLOGICAL FRAMEWORK

The purpose of the study is to establish the features of the dynamics of interest in the Big Data topic in the US and Russia regarding global trends through collection via Google Trends and subsequent regression analysis of relevant analytical data.

Using the Google Trends data (Google Trends), the following tasks were solved:

- The construction of regression models of the correlation of interest in the Big Data topic in the US and Russia with respect to global trends;
- The construction of time series characteristics of the frequency of queries for the phrase Big Data in the Google search system in the United States and Russia in 2012-2017;
- The construction of time series for changing characteristic $K_{\text{interest rate growth}}$ in relation to the Big Data topic in the US and Russia in 2012-2017;
- Establishment of a pattern of the correlation of the level of interest in the Big Data in Russia in 2012 – 2017 in relation to the analogue in the USA;
- Formulation of conclusions on trends and patterns of public interest in the Big Data in the world, the United States and Russia.

RESULTS AND DISCUSSION

Figure 2 represents a diagram of the change in the frequency of queries of the phrase "Big Data" in the Google search engine in the world in the period from January, 2012 to September, 2017 (Google Trends).

Figure 2: Screenshot of the screen mode of Google Trends to identify the frequency of Big Data requests in Google in the world: A – time series data on the frequency of Big Data queries in Google; B – large scale of data
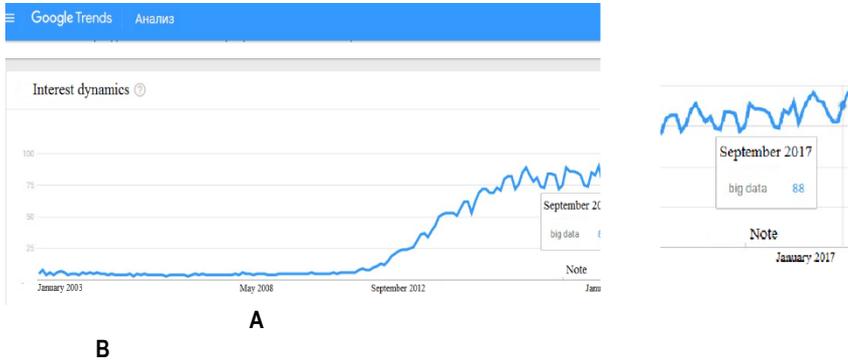
Based on similar data obtained for the US and Russia, regression models (Borovikov, 2003) were constructed that identify the correlation of interest in the Big Data topic in the US and Russia with respect to global trends (Figure 3).

Figure 3: Regression models identifying the correlation of interest in the Big Data topic in the Google search system in the US and Russia regarding global trends
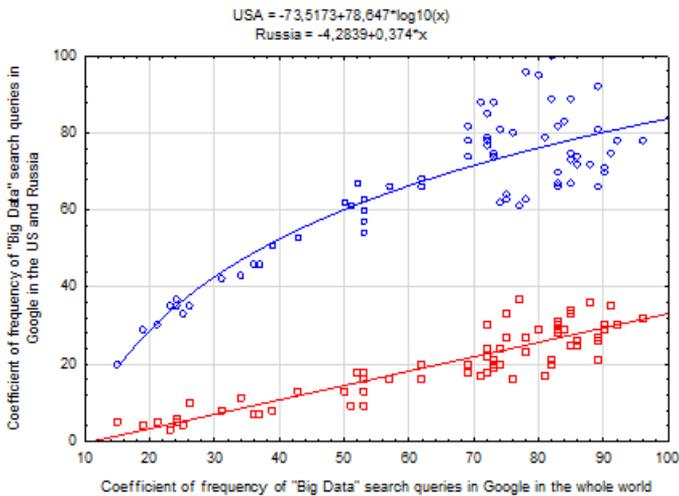
The models shown in Figure 3 show that the interest in the Big Data topic in the US is much higher than the analogues typical for Russia. However, there are two important features of the situation.

- At the initial stage, the frequency of querying Google for "Big Data" data in the United States exceeded the frequency of requests in the world (in 2012-2014), but since a certain time (2015), this trend ended, the situation leveled off, and today there are already signs of a turnaround.
- For Russia, there is a slow but steady increase in interest in the Big Data topic, although relative to the global level, the coefficient of interest in the topic in Russia is now equal to 0.3.

PETROV et al.
Trends and Patterns of Public Interest to Big Data in International Community
390

Comparative time series of the interest characteristic of the Big Data topic in Google for the US and Russia are presented in Figure 4. Regarding the maximum interest (100 points) to the Big Data topic, which was shown in the US in October 2014, today there is a drop in the number of search requests on the Internet by about 20 – 30%. In Russia, interest in this topic is continuously growing, and the growth rate of such interest, expressed by the coefficient $K_{interest\,rate\,growth}$ (1), is already significantly higher than the analog for the United States (Figure 5).

Figure 4: Time series characteristics of frequency of queries for the phrase Big Data in the Google search system in the US and Russia in 2012 – 2017
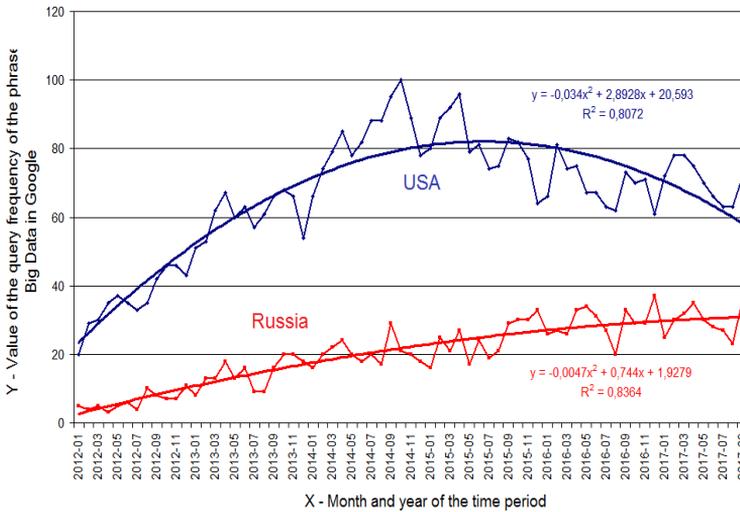
$$K_{interest\,rate\,growth} = \frac{N_{actual\,month}}{N_{01.\,2012}}$$

(1)

Where:

$K_{interest\,rate\,growth}$ – coefficient of interest growth rate to the topic in a certain period of time;

$N_{actual\,month}$ – the level of interest in the Google search engine to the topic, taking into account the highest score for a certain region and the time period. 100 points mean the highest level of popularity of the request;

$N_{01.\,2012}$ – the level of interest to the topic in the Google search engine in January, 2012 (the initial period of tracking the frequency of user requests).

Utopía y Praxis Latinoamericana; ISSN 1315-5216; ISSN-e 2477-9555
Año 23, n° 82 (julio-septiembre), 2018, pp. 386-394

391

Figure 5: Time series of the change in characteristic in relation to the Big Data topic in the USA and Russia in 2012-2017
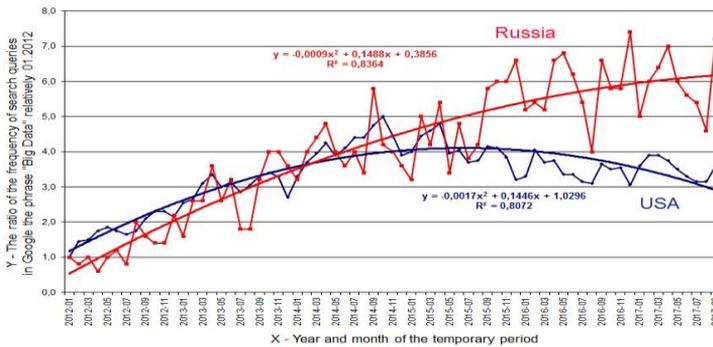
By the middle of 2014, the value of $K_{\text{interest rate growth}}$ in Russia has reached the value of an analogue typical for the USA (about 3.5); in subsequent years, this indicator in Russia exceeds the analogue for the US and today there is already a two-fold gap between Russia and the US on this indicator (6 in Russia against 3 in the US). The graph in Figure 5 really shows that in the US, interest in the Big Data topic began to decline around the spring of 2015, which is in line with Gartner's data (Figure 1) (the phase of getting rid of the illusions *Trough of Disillusionment* on the Hype Cycle chart).

The model of changing the correlation of the level of interest in the Big Data in Russia in 2012-2017 in relation to the analogue in the USA is also interesting (Figure 6).

Figure 6: Model of correlation of the level of interest in the Big Data topic in Russia in 2012-2017 in relation to the analogue in the USA

PETROV et al.
Trends and Patterns of Public Interest to Big Data in International Community
392
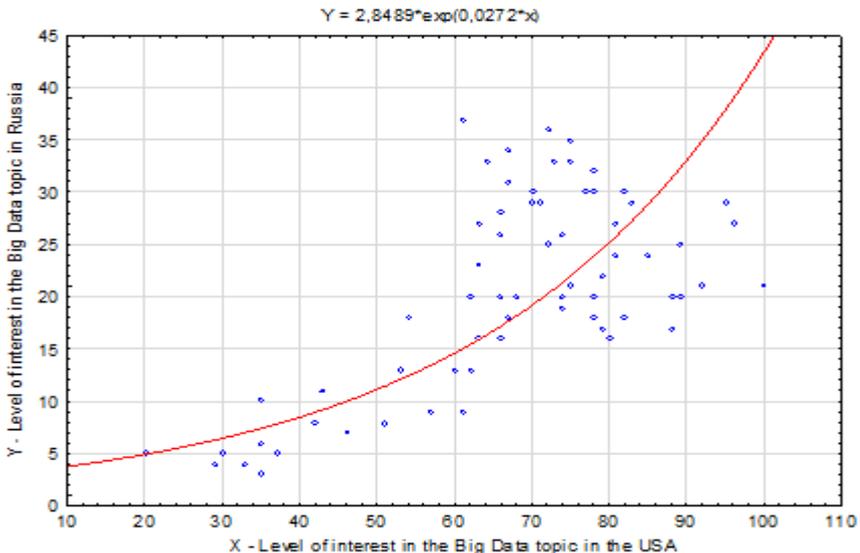
Formally, this model is exponential. The coefficient of determination of the model $R^2 = 0.448$ is relatively low, although the value of the Fisher test F = 54.27 exceeds the tabulated one, which allows us to speak about the validity of using an exponential model for describing the process under study. Visual analysis shows that at the level of interest to the topic in the US = 70 and interest in the topic in Russia = 20 (approximately at the beginning of 2014), the model ceases to be exponential.

In order to find an explanation for this phenomenon, a spline plot was constructed (Figure 7). Spline is a function that, together with several derivatives, is continuous on the whole given interval [*a, b*], and on each partial interval [x_i, x_{i+1}] is an algebraic polynomial. In fact, the spline shows the trajectory of the process under study in the phase space (Borovikov, 2003).

Figure 7: Spline graph of the ratio of the level of interest to the Big Data topic in Russia in 2012 – 2017 in relation to the analogue in the USA: A – initial, for the entire period 01.2012-09.2017; B – final, two phases of the process of changing interest in the Big Data topic in Russia

A

B
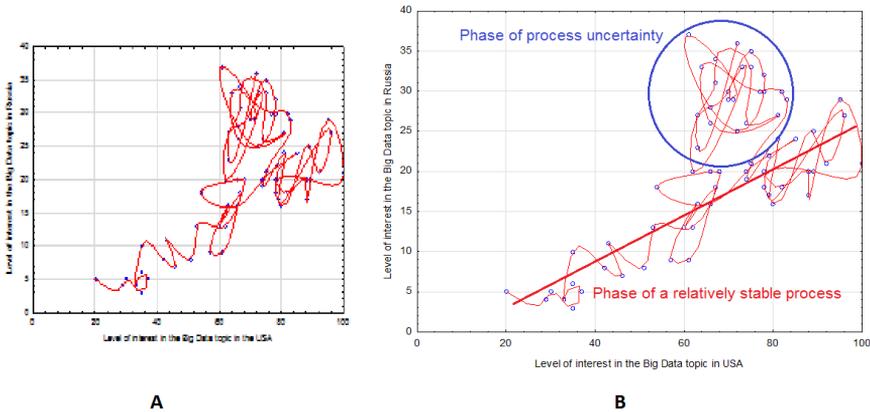
Analysis of the spline graph shows that up to a certain point the process developed relatively stable (phase 1), but over time the process passed to phase 2, which can be described as the "phase of process uncertainty". Despite the seemingly steady growth of interest in the Big Data topic in Russia (Figure 4) and even relatively higher growth rates in Russia than in the US (Figure 5), the phase portrait of the process shows that in Russia today there are times of loss of understanding of Big Data technology benefits and there comes a stage of stagnation. After this, it is quite possible that there will be a degradation of interest in this topic in Russia (as has already happened in the US, as noted by Gartner experts).

CONCLUSION

Based on the results of the study, the following conclusions can be drawn.

1.   In general, the assumption formulated as "with a generally relatively lower level of development of Big Data technologies in Russia at the moment, interest in this technology in our country is higher than in the US and the world" has good reason to be considered objectively correct. An important note is that this trend is supposedly coming to an end.

2.   Regression models of the correlation of interest in the Big Data in the US and Russia with respect to global trends have shown that Russia loses significantly in this interest to both the world community

Utopía y Praxis Latinoamericana; ISSN 1315-5216; ISSN-e 2477-9555
Año 23, n° 82 (julio-septiembre), 2018, pp. 386-394

393

and the US, but Russia has a steady growth of this interest, while interest in the subject in the US decreases;

3. In terms of growth of interest in the Big Data in 2012 – 2017, Russia caught up and bypassed the US (approximately in mid-2014). In the fall of 2017, the growth rate of interest in the topic in Russia is about twice as high as in the US;

4. An analysis of the exponential model of the growth of interest in the Big Data topic in Russia regarding the US trend by plotting the spline showed that approximately from the end of 2015 the Russian trend ceased to be steadily growing and moved to a new phase – the uncertainty of the process.

5. In general, it can be stated that the public interest in the Big Data topic in Russia repeats trends and patterns typical for the US, but with a time lag. This explains why there is a relatively low level of interest in the topic in absolute terms and a higher rate of growth of interest in the topic in Russia regarding the US.

BIBLIOGRAPHY REFERENCES

Agneeswaran, V. (2012). Big-data – theoretical, engineering and analytics perspective, in: S. Srinivasa, V. Bhatnagar, eds. *International conference on Big Data analytics,* pp. 8-15, Springer, Berlin, Heidelberg.

Avdeeva, I.L. (2016). Analysis of foreign experience using global technologies "Big Data", *Naukovedenie.* 8(6): 1-11. Retrieved June 1, 2018 from http://naukovedenie.ru/PDF/13EVN616.pdf

Bhadani, A., Jothimani, D. (2016). Big data: Challenges, opportunities and realities, in: M.K. Singh, D.G. Kumar, eds. *Effective Big Data management and opportunities for implementation,* pp. 1-24, IGI Global, Pennsylvania.

Borovikov, V.P. (2003). *STATISTICA. The art of data analysis on a computer: for professionals*. Piter, St. Petersburg.

Boyd, D., Crawford, K. (2012). Critical questions for big data, *Information, Communication & Society.* 15(5): pp. 662-679.

Campbell, P., ed. (2008): *Nature,* Vol. 455, no. 7209. Retrieved from May25, 2018 from https://www.nature.com/search?q=7209&order=relevance&journal=nature

Chernyak, L. (2011). Big data – a new theory and practice, *Open systems. The database management system*. 10: p. 18.

Cukier, K. (2010). Data, data everywhere, *The Economist.* Retrieved June 15, 2018 from http://www.economist.com/node/15557443

Gartner. (2015). *Gartner's 2015 hype cycle for emerging technologies identifies the computing innovations that organizations should monitor 2015 hype cycle special report illustrates the market excitement, maturity and benefit of more than 2,000 technologies*. Retrieved June 15, 2018 from https://www.gartner.com/newsroom/id/3114217

Google Trends. *Analysis*. Retrieved June 15, 2018 from https://trends.google.ru/trends/explore?date=all&q=Big%20Data

Jin, X., Wah, B.W., Cheng, X., Wang, Y. (2015). Significance and challenges of big data research, *Big Data Research.* 2: pp. 59-64.

Li, H., Lu, X. (2014). Challenges and trends of big data analytics, in: J.E. Guerrero, ed. *Proceedings of ninth international conference on P2P, parallel, grid, cloud and internet computing (3PGCIC)*, pp. 566-567, IEEE, Danvers, MA.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H. (2011). Big data: The next frontier for innovation, competition, and productivity, *McKinsey Global Institute*. Retrieved June 15, 2018 from https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Meyer-Shenberger, V. (2014). *Big Data. A revolution that will change the way we live, work and think*. Mann, Ivanov and Ferber, Moscow.

Miller, H. G., Mork, P. (2013). From data to decisions: A value chain for big data, *IT Professional*. 15(1): pp. 57-59.

Von Grassegger H., Krogerus M. (2016). Ich habe nur gezeigt, dass es die Bombe gibt. Der Psychologe Michal Kosinski hat eine Methode entwickelt, um Menschen anhand ihres Verhaltens auf Facebook minutiös zu analysieren. Und verhalf so Donald Trump mit zum Sieg. *Das Magazin.* Retrieved June 15, 2018 from https://www.dasmagazin.ch/2016/12/03/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt/