

API-Sociology and Google Global society: confluence of factors for political and economic awareness

Lev Mazelis

Department of Mathematics and Modeling
Vladivostok State University of Economics and Service
Vladivostok, Russia
lev.mazelis@vvsu.ru

Andrey Mazelis

Department of Mathematics and Modeling
Vladivostok State University of Economics and Service
Vladivostok, Russia
Andrey.Mazelis@vvsu.ru

Nina Shamshurina

Institute of sociology, psychology and humanities
I.M. Sechenov First Moscow State Medical University
Moscow, Russia
shamshuriny2@mail.ru

Yuriy Kolesnichenko

Uzgraph.ru
Moscow, Russia
green-apple_2000@mtu-net.ru

Dariya Yakovleva

Department of Mathematics and Modeling
Vladivostok State University of Economics and Service
Vladivostok, Russia
darya.yakovleva15@vvsu.ru

Olga Kolesnichenko

Institute of sociology, psychology and humanities
I.M. Sechenov First Moscow State Medical University
Moscow, Russia
oykolesnichenko@list.ru

Anna Zykova

Institute of sociology, psychology and humanities
I.M. Sechenov First Moscow State Medical University
Moscow, Russia
a.zykova@mgnot.ru

Ivan Grigorevsky

Aylamazyan Program Systems Institute of RAS
National Supercomputing Technology Platform
Pereslavl Zalessky, Russia
gin@nscf.ru

Abstract — Text Big Data Analytics Study is presented (frequency-morphological analysis of open texts from Google). Visualization of Morphological Matrix, cluster analysis (hierarchical method of dendrograms creation), and Value Chain Map were implemented as the set of tools for analytics. World Economic Forum's Global Risks Report was taken as a model in studying global issues. Global Trends Map was created and is presented; it indicates global social persistent inequality. Our results were compared with Global Risks Report and Multiple Futures Project–2030. Value Chain Map, based on Text Big Data Analytics for eight countries, helped to assess economic situation in terms of competitiveness at the global market.

Keywords — Text Big Data Analytics, Google, API-sociology, Data Mining, Cluster Analysis, Fourth Industrial Revolution

I. INTRODUCTION

Text Big Data Analytics study has continued during four years. Ad-hoc Big Data Analytics Team has been established among universities in uniting framework of National Supercomputing Technology Platform – National Supercomputing Forum. This type of approach consolidates

students and researchers on the field of scientific interest. Academic ad-hoc approach allows to gain the new skills and experience and helps in professional knowledge growth.

Data Mining conditions were as following. Unstructured arrays of keywords from open Google resources have been investigated thorough API access for understanding of global society processes. Selected keywords were counted using *special software*, which has been created at the Vladivostok State University of Economics and Service. This is the frequency-morphological analysis of open texts. Morphological Matrix of 52 keywords was collected and is presented in Table 1. Keywords counted in million, in conjunction with 100 countries names and 2015-2016 years of texts publication on the Internet.

The previous results of our study showed that frequency of selected keywords has correlation with various social and economic trends [1], [2]. The conceptual scheme of analytics was creating step by step. The previous Morphological Matrix after Data Mining and collecting was transformed into Ontology of metadata. *At this stage of study* the Ontology has

been expanded; visualization of Morphological Matrix, cluster analysis k-means and hierarchical method of dendrograms creation, ranking and Value Chain Map were implemented as a set of tools for analytics.

This study can be called API-sociology. Many keywords about many issues can be investigated. Internet is the reflection of different complex processes from level of every human to local countries' level and global world level. *The aim for researchers* is to find the hidden patterns and meanings in frequency distribution of selected keywords.

TABLE I. MORPHOLOGICAL MATRIX

Units	Keywords
Unit Information	computer, cloud computing, mobile phone, dollar exchange rate, euro exchange rate, mobile app, software, artificial intelligence, smartphone
Unit Energy & Material	electric cars, solar panel, drip irrigation, gas supplies, oil, oil price, nuclear power plant, 3D printing, robotics, GMO, GMO harmful
Unit Social inequality	crisis, inflation, price increase, unemployment, dismissal, poverty
Unit Stress load	migrants, refugees, Muslims, terrorism, terrorist, occupation, narcotic, alcoholism, morbidity, mortality, violation, crime, war, casualties
Unit Social profile	revolution, demonstration, protest, strike, corruption, stability, prosperity, democracy, development, freedom, human rights, justice

II. RESEARCH TOPICS AND GLOBAL TRENDS MAP

A. Study methods of Analytics (tools)

The expanded Ontology of metadata includes several analytical approaches with qualitative and quantitative methods. Global Trends Map (Visualization of Model of Google's most discussed keywords) and Value Chain Map were created in Microsoft Excel; both as qualitative technique help to provide insights into the research issues. For dendrogram analysis the nearest neighbour clustering was implemented (R, MATLAB, Python); this quantitative technique classifies data into homogeneous subgroups that also provides insights.

B. Global Discussion Landscape

This study is dedicated to investigation of global issues. Could the elaborated set of analytics tools and Morphological Matrix be used for assessment of global discussion landscape? Is it possible to understand the future human development risks? The World Economic Forum's annual Global Risks Report (GRR) [3] was taken as a model in studying global issues. In contradistinction to the Text Big Data Analytics, GRR is the classical sociological survey that includes around 900 respondents (experts). Survey's respondents have been asked to determine the most important trends in shaping global development.

The Global Risks Landscape and Risks-Trends Interconnections Map are presented in GRR; the scalable markers indicate five spheres: economic, geopolitical, environmental, societal, and technological. In our study the same spheres are referred in the five units of keywords: Unit Information (modern technology and financial resources); Unit Energy & Material (modern and basic technology; GMO harmful as environmental problem); Unit Social inequality (economic and social issues); Unit Stress load (geopolitical and social issues); Unit Social profile (institutional systems, social and economic issues).

GRR-2018 highlights the several main concerns: social persistent inequality and unfairness, domestic and international political tensions, environmental dangers, and cyber vulnerabilities. The trend of profound social instability is in the spotlight of GRR respondents; it includes social movements, protests, and social unrest, which disrupt stability, negatively impacting populations and economic activity. Other trends are chronic economic problems, related to earnings, unemployment, and inequality.

The current economic transformation is called Fourth Industrial Revolution. Rapid technological change increases unemployment rate as adverse impact of technological advance. Mostly it concerns 3D printing in industry, Artificial Intelligence and robotics. Information Technologies as Cloud Computing, mobile communication, mobile services, software, and computers development are prevailing today. In parallel, attention is focused on oil and gas.

Based on the Text Big Data Analytics and Morphological Matrix analysis the Model of Google's most discussed keywords – *Global Trends Map* was created in our study; it is presented at Fig. 1. The Global Trends Map was built using the most frequent keywords for each of the 100 countries. Peaks of the biggest patterns are shown. The closer to center on diagram, the more number of countries have a keyword with high frequency (all 100 countries, or near 100). The size of the marker is scalable and also conditionally reflects the number of countries, which have a keyword with high frequency. There are four leading positions on the Global Trends Map: mobile phone, development, revolution, poverty. These keywords reflect the interest in Information Technology and economic development. Global Trends Map indicates the topicality of *social instability and chronic economic problems* (keywords "revolution" and "poverty"). Some part of the patterns of keyword "revolution" can be associated with Fourth Industrial Revolution.

The markers with less size are shown farther from the center on the Global Trends Map, i.e. less number of countries, which have a keyword with high frequency (half of countries). These are additionally signs of social instability and chronic economic problems: protest, strike, demonstration, occupation (job or profession), and also deep social problems as morbidity, narcotic (drug addiction).

Around the periphery even less markers are located, i.e. even less number of countries, which have a keyword with high frequency. But it is still the trend for some countries (less than 30). The social instability signs are added to common trend: crisis, price increase, unemployment. Keywords

“alcoholism” and “crime” reflect deep social problems. Another keyword related to social issue is “democracy”. Altogether the Global Trends Map in our study indicates social persistent inequality and unfairness as it has been highlighted in GRR.

Domestic and international political tension also is identified on the Global Trends Map; there are some related keywords: Muslims, terrorism, terrorist, war. In this study the environmental dangers and cyber vulnerabilities were not investigated using keywords. There are some keywords around the periphery on the Global Trends Map, which reflect critical importance of *Information Technology* (computer, software, mobile app, smartphone). The keywords “artificial intelligence”, “3D printing” and “robotics” haven’t yet reached the peak level within Morphological Matrix. These technologies still proliferate at implementation stage. They together constitute the Fourth Industrial Revolution, which will disrupt previous social order and stability.

Also the signs of *old technological order* are presented on the Global Trends Map: gas supplies, oil, oil price. It is important to mention that the hydrocarbons during Fourth Industrial Revolution will change the function from fuel to printing ink for 3D-printing mass consumption products. The new uprise of hydrocarbons market is expected in near future.

The two keywords “GMO” and “GMO harmful” constitute peripheral trends. These keywords reflect the global discussion about *food crisis* and measures how to overcome food crisis with less damage for people's well-being and health. According to the GRR world will face insufficiency of global food supply, including reduced crop diversity, crop diseases and climate change. Marker of food crises was around the periphery on the GRR Risks-Trends Interconnections Map, as well as keywords “GMO” and “GMO harmful” are located at the periphery on the Global Trends Map in our study.

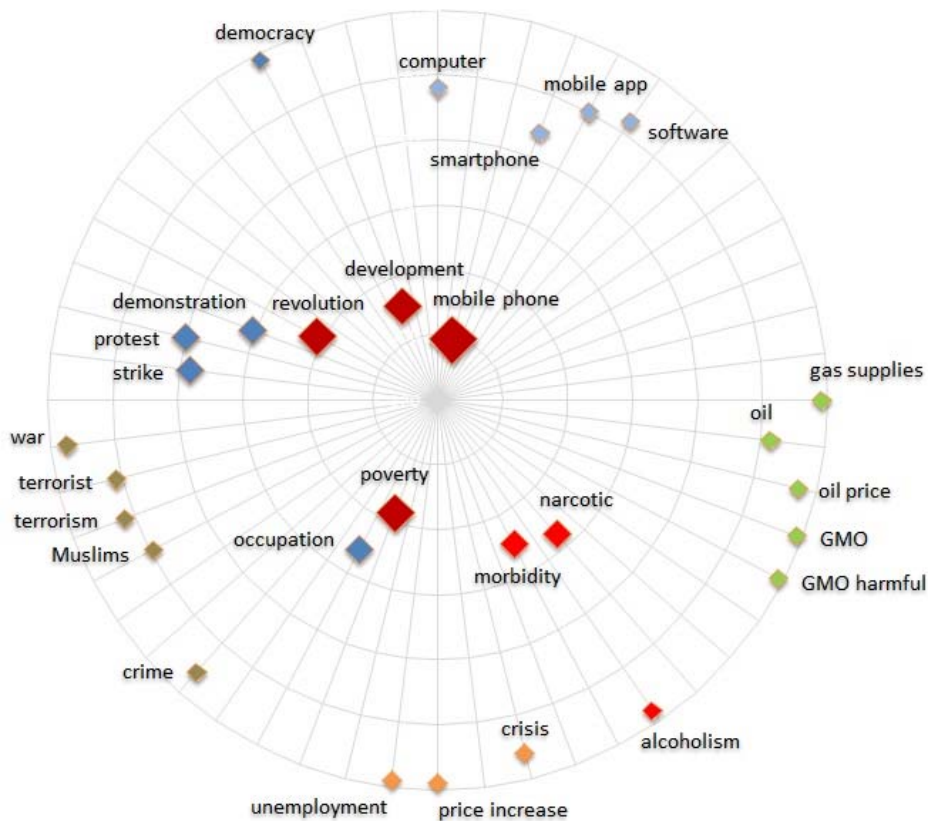


Fig. 1. Global Trends Map created due to API-sociology approach. Morphological Matrix analysis, 100 countries, 2015-2016 years of texts publication in Google. See the description in text.

C. Predictive analytics and indicators without imposed survey condition

Multiple Futures Project – Navigating towards 2030 Report (MFP) [4] was another big international survey, which was taken into account for our results consideration. MFP is the one-off issue. It aimed at understanding of the future

threats’ environment through rigorous analysis of the emerging challenges. MFP is based on many workshops and experts’ opinions from 45 countries and more than 60 institutions working in political, military, civil and economic fields. During experts brainstorm the main question was asked: what are the threats and challenges that will pose the greatest risk to the interests, values and populations in future?

Four futures were described as a result of debate regarding the risks and vulnerabilities that endanger the population.

- Dark side of exclusivity. Social scenario. This future will be complicated by social instability, poverty, challenges to state authority, demographic pressure, and deteriorating environmental conditions.
- Deceptive stability. Criminal scenario. For this future the following risks have been identified: transnational criminal movements; potential spill-over of ethnic, religious, and ideological conflict; terrorism; uncontrolled migration.
- Clash of modernities. Technological scenario. In this future in parallel with strong belief in rationalism coupled with technological innovation will be the following risks: internal tension between technocratic efficiency and civil liberties; exploitation of technologies by criminal / rogue elements; vulnerability of complex interdependent systems and infrastructures. This scenario is more relevant to the time when Fourth Industrial Revolution will be completed.
- New power politics. Military scenario. This future will characterize by competition for ideological supremacy, competition for resources, proliferation of nuclear and other weapons of mass destruction, interstate rivalry and war.

Experts of MFP, which work in different spheres including military, had the prescribed, imposed condition for discussion during workshops / seminars. They described all known global problems and finally the four conclusions were given as multiple options. MFP has special emphasis on terrorism and security issues (predetermined result because of NATO involvement). In GRR the trends were determined due to imposed survey condition in one direction – economy & business. Therefore, the GRR conclusion and findings are mainly focused on economy (predetermined result because of World Economic Forum involvement). But is it really right and precise global situation awareness, or is it only possible assessment by experts related to their interests? The assessment of possible scenarios as abstract reasoning is important for forecast, but it does not give awareness and actual accent about what is really going on.

In our study of *API-sociology* there were no survey or workshops, no imposed questions or moderators, and the findings were not predetermined. The frequency-morphological analysis of open texts from the Internet (Google) allows taking into consideration the maximum possible coverage of people and opinions, and all three levels of Global society behavior on the Internet:

1) *Activity*. It depends on number of IP connections, Internet accessibility in every country, and amount of country's population;

2) *Knowledge*. Internet's open textual resources reflect the different features of technological evolution in countries and what do people know about different technology;

3) *Involvement*. People write more on the Internet about that problems which concern them more.

API-sociology helps to understand influence of Globalization on local scale that is called Glocalization process. Altogether API-sociology approach provides more flexible and sensitive tools for global situation awareness. As the result the Model of Google's most discussed keywords was built. Global Trends Map shows the picture which reflects the real, not predetermined, dominant, prevailing topics. The picture is closer to GRR findings and related to MFP *Social scenario* Dark side of exclusivity. The themes of terrorism, war, Muslims and crime have less relevance for world discussion in compare with MFP four scenarios.

Comparison of the API-sociology study findings with two reports' conclusions based on survey or workshops, makes it possible to understand the differences. API-sociology findings show the real prevailing spectrum of interests and concerns of the global audience, but experts' forecast reflect only their views and concerns. Our recommendations can be follows: every management decision, made on basis of such reports as GRR or MFP, should be accompanied by API-sociology global assessment. Only this combination of approaches will give the comprehensive political and economic awareness.

III. DENDROGRAM ANALYSIS AND VALUE CHAIN MAP

A. Dendrogram analysis (hierarchical clustering or tree diagram)

In this paper we present the dendrogram for the first time in our four-year study. At Fig. 2 the dendrogram (fragment) shows the two big clusters. The single-linkage clustering (nearest neighbour clustering) was implemented. The vertical axis of dendrogram represents the distance or dissimilarity between clusters. The horizontal axis represents the objects (keywords, its frequency) and clusters.

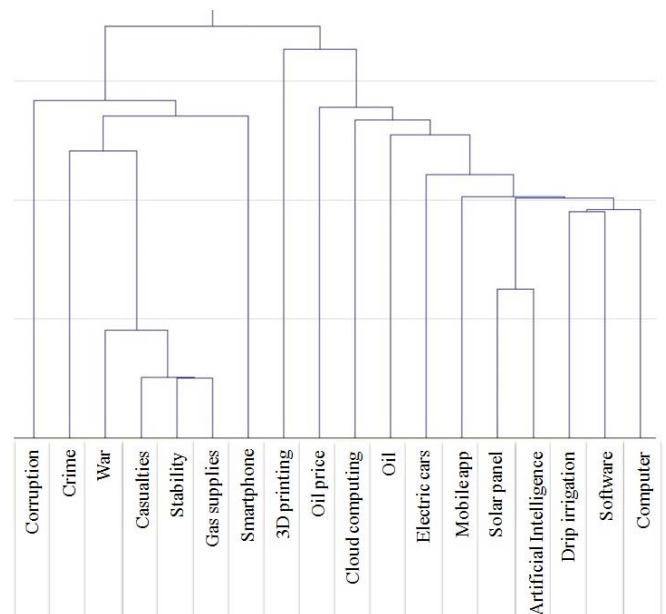


Fig. 2. The fragment of the dendrogram of Morphological Matrix, 52 keywords and 100 countries, 2016; analyzed parameters – keywords.

Right cluster is characterized by more homogeneous distance and it is of interest for description. Group of keywords is united by one narrative – Fourth Industrial Revolution. The keywords of this cluster within Morphological Matrix of all 100 countries have the similar and related frequency distribution for each country. Hydrocarbons (keywords “oil” and “oil price”) are changing the function from fuel to printing ink for 3D-printing mass consumption products, including full-body 3D-printed cars

made by carbon fiber reinforced plastic. Artificial Intelligence is now most prominent as fully autonomous, self-driving car with computer and software. Cars are connected to Internet of Things framework due to Cloud computing and Mobile apps. Solar panels stations are building around the world for electric cars’ battery charging. Presented on the dendrogram the relationship between all mentioned topics, united into one narrative within the right cluster, is reflected *the new automotive industry*.

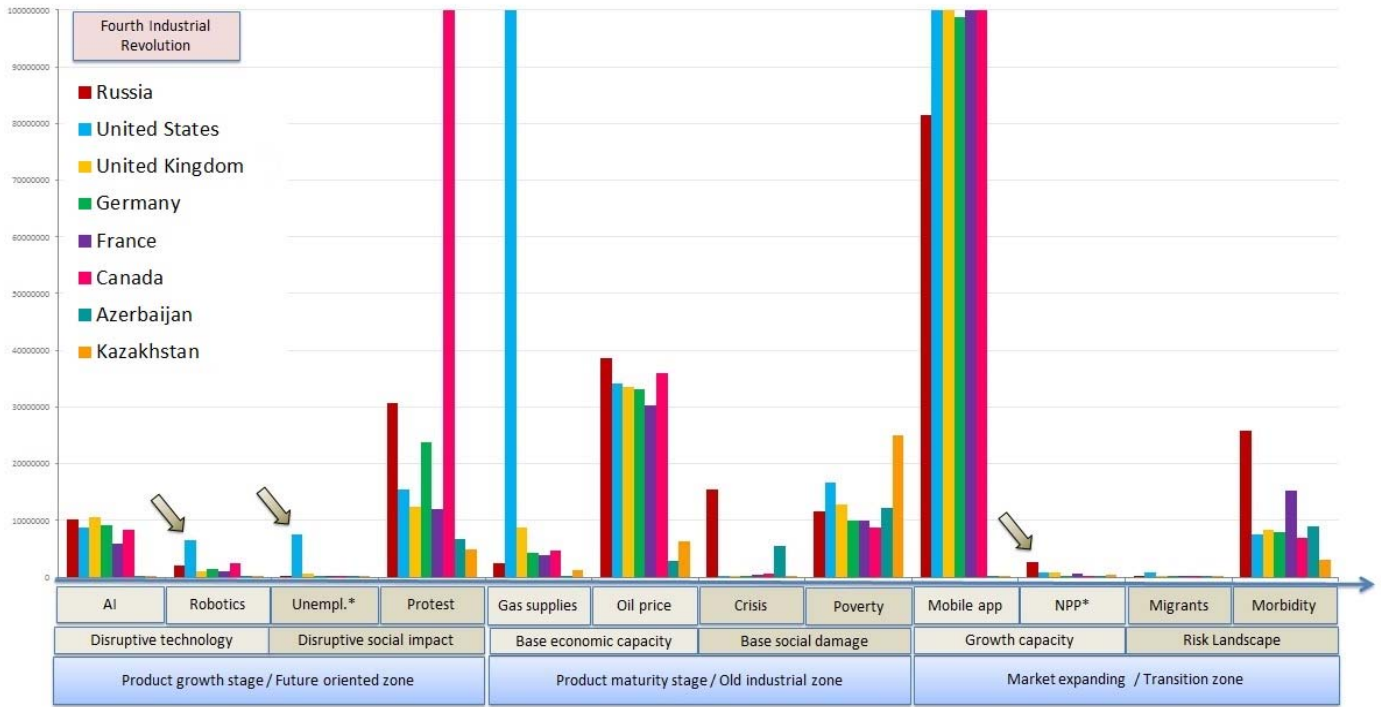


Fig. 3. The modified Value Chain Map based on the results of Text Big Data Analytics from Google (frequency-morphological analysis). The number of keywords in million during 2015-2016 in conjunction with the countries names. All the components of the Value Chain Map are evolving from left to right due to demand competition. Level above 100 million isn't shown. *Abbreviation: Unempl. – unemployment; NPP – Nuclear power plant.

B. Value Chain Map

For analysis of the new Technological Order accepting by some different countries [5], [6], [7] the special *Value Chain Map* was elaborated. This approach has been presented at the previous stage of our study [1]. At Fig. 3 the Value Chain Map is presented for eight countries: Russia, United States, United Kingdom, Germany, France, Canada, Azerbaijan, and Kazakhstan.

Value Chain Map helps to reveal hidden meanings in results of Text Big Data Analytics. We can assess the economic competitiveness and the shaping of future consumer market, looking to patterns of counted keywords in dynamics. Value Chain Map is evolving from left to right due to demand competition.

Product growth stage (Future oriented zone) is related to early beginning of the new Technological Order (Fourth Industrial Revolution). This is the most competitive stage of industry with Artificial Intelligence and robotics, both are

disruptive technologies, which cause disruptive social impact. The United States show the increased interest in robotics, and at the same time the growth of theme of unemployment is shown on Value Chain Map. This bound rise of two patterns correlates with the fact that the United States keep on leadership and now are ahead of other countries in the new Technological Order development, simultaneously undergoing equivalent disruptive social impact. Other countries should expect the same social impact in the future.

Product maturity stage (Old industrial zone) is related to previous century. For many countries hydrocarbons is still base of economy. But it can be noticed that interest in hydrocarbons now is less than interest in digital economy drivers of growth – Mobile applications. The United States show the increased interest in gas supplies, that can be explained by emerged market of carbon materials as printing ink for 3D-printing. Hydrocarbons will change the role from fuel to new materials.

Market expanding (Transition zone) is associated with digital economy of 21st century, widely distributed market of Cloud computing, Internet of Things, and Mobile applications. Looking to Value Chain Map's data for Azerbaijan and Kazakhstan can be recommended pay more attention to Mobile applications. Alongside with Mobile apps, the electricity energy supply also constitutes the critical condition for Internet of Things' market expansion [7], [8]. Russia focuses on nuclear energy that is shown on Value Chain Map (predominant number of counted keywords "Nuclear power plant").

Digital economy means the economy of human capital that requires increased attention to people's health. Russia shows increased level of counted keywords "morbidity". This is sign of implemented the state's strategy during recent years to enhance health monitoring and extend lifetime of Russia's population [9].

Thus Value Chain Map is the efficient additional method for Text Big Data Analytics. It shows the detailed information about countries' prospects and competition.

IV. CONCLUSION

Big Data brings fundamental changes in sociological studies. Surveys and Focus group method were developed over many decades, and are used for preparing many global reports. Such reports as The World Economic Forum's annual Global Risks Report and Multiple Futures Project – Navigating towards 2030 Report are based on classical sociological approaches. Big Data is creating the significant new opportunity to reveal new value and new understanding [10]. Surveys and Big Data are complementary data sources, they are not competing. There are differences between the approaches, but it should be seen as an advantage rather than a disadvantage [11].

Implementing the current stage of our four-year study we have achieved following results.

- Due to special software, which has been created at the Vladivostok State University of Economics and Service, we collected the unique Morphological Matrix of 52 keywords for 100 countries from open Google resources. This type of study we named API-sociology; this is the term that we introduced into sociology during four years as the research at the junction of social Big Data & mathematical modeling. We believe that the developed Morphological Matrix will be used by other researchers.
- The Model of Google's most discussed keywords – Global Trends Map was created in our study. It allows us to reveal the most discussed problems; Global Trends Map indicates the topicality of social instability and chronic economic problems in the world.
- For the first time in our four-year study we present the dendrogram analysis as the tool for Text Big Data

Analytics. Dendrogram analysis shows how keywords are united by one narrative. We revealed cluster, which reflects the shaping of the new automotive industry in the world.

- Value Chain Map allows putting keywords in needed sequence to determine what has already covered previous stage of development, what will be the innovative stage, what positions countries have already taken at the global market, and what countries have the advantages. Presented Value Chain Map shows the negative influence of Fourth Industrial Revolution to employment, and prevalence of interest in digital economy.

We have elaborated API-Sociology step by step. The confluence of factors related to Internet global audience and chosen tools for analysis allows gaining political and economic awareness in advance. It helps in strategic planning and decision-making process. We recommend using surveys / workshops method and API-Sociology methods together. Presented set of tools can be elaborated as common platform for predictive Text Big Data analytics.

REFERENCES

- [1] O. Kolesnichenko, G. Smorodin, D. Yakovleva, L. Mazelis, O. Zhurenkov, and Y. Kolesnichenko, "Text Big Data Analytics: exploring API opportunity". 10th IEEE International Conference on Application of Information and Communication Technologies. Qafqaz University, Azerbaijan, 2016, pp. 37-41.
- [2] O. Kolesnichenko, G. Smorodin, D. Yakovleva, L. Mazelis, S. Balandin, and Y. Kolesnichenko, "Index of Intrastate Stability prepared by Text Big Data analytics". 7th International Conference on Cloud Computing, Data Science & Engineering – Confluence. Amity University, India, 2017, pp. 182-187.
- [3] The Global Risks Report, 13th Ed. World Economic Forum, Geneva, 2018, 72 p.
- [4] Multiple Futures Project – Navigating towards 2030, Report of ACT NATO, 2009, 69 p.
- [5] Kondratiev's waves: Heritage and Modernity. Ed. by L.E. Grinin, A.V. Korotaev, and V.M. Bondarenko. Volgograd, Russia: Uchitel, 2015, 256 p.
- [6] E. Toffler, The Third Wave. Moscow, Russia: AST, 2004, 781 p.
- [7] J. Manyika, S. Ramaswamy, S. Khanna, H. Sarrazin, G. Pinkus, G. Sethupathy, and A. Yaffe, "Digital America: A tale of the haves and havemores". Report, McKinsey Global Institute. United States, 2015, 107 p.
- [8] O. Vermesan and P. Friess, "Digitising the Industry. Internet of Things Connecting the Physical, Digital and Virtual Worlds". River Publishers Series in Communications, Vol. 49. Denmark-Netherlands, 2016, 339 p.
- [9] Prospect for the socioeconomic development of the Russian Federation: analytical report. Ed. by N.N. Kalmykov. Russian Presidential Academy of National Economy and Public Administration, Russia, 2017, 452 p.
- [10] D. Dietrich, B. Heller, and B. Yang, Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, USA: Wiley, 2015, 410 p.
- [11] L. Japiec, F. Kreuter, M. Berg, P. Biemer, P. Decker, C. Lampe, J. Lane, C. O'Neil, and A. Usher, AAPOR Report on Big Data. The American Association for Public Opinion Research, USA, 2015, 50 p.