Technology of Application of Software Tools for Energy Technology Forecasting

Alex Kopaygorodsky [1,2]
[1]Melentiev Energy Systems Institute of Siberian Branch
of the Russian Academy of Sciences
[2]Irkutsk National Research Technical University
Irkutsk, Russia
kopaygorodsky@isem.irk.ru

Abstract—In this paper an approach to the collection and analysis of information for the task of scientific and technological forecasting is considered. The approach is based on the use of ontologies, Tech Mining and Data Analysis methods. The developed approach is supported by the distributed intellectual information system created by the author. The technology of using tools for solving the problem is presented.

Keywords—intelligent system, research support, Big Data, semantic analysis.

I. INTRODUCTION

Over the past decade there has been an active development of a large number of intelligent approaches and methods to provide decision support for planning and managing innovative economic development [1-3]. The use of Data Science methods in high-tech industries can significantly increase the effectiveness of managerial decisions. A similar approach to management in individual organizations at various levels is commonly called "data-based management". The main source for making managerial decisions is the analysis of information and data collected from various sources to identify existing and emerging trends and reduce the reaction time to them, which reduces losses, increases profits and ensures the sustainable development of the company. The results of scientific and technological forecasting of the development of individual industries allow us to pay more attention to promising research, which allows us to save time and money.

II. PREDICTION BASED ON DATA ANALYSIS

Despite the fact that in the United States and the UK, sound scientific and technological forecasts and innovative development programs are one of the most important established forms of economic regulation, but over the past few years, active research has been carried out by scientific teams from China. The main sponsor of Chinese researchers is the National Natural Science Foundation of China (NSFC). Such an interest is primarily due to the rapid development of information technology and the onset of the Big Data Age. When implementing this approach to management in commercial organizations, specialized information and analytical departments are created under the direction of the Chief Digital Officer / Chief Data Officer (CDO). The opinion of CDO is the key in the development of the company, identifying new business opportunities, providing access to new market segments, introducing fundamentally new products, services to the market, etc. Ignoring technical and economic trends can be fatal or cause significant financial damage. Examples of such companies are Polaroid, Kodak, Motorola, 3Com, Sun Microsystems. The application of the described approach is relevant in solving the problems of forecasting and organizing monitoring of innovative technological solutions in any sector of the Russian economy, and especially for the development of energy infrastructure. Energy is the basic industry for the functioning of other industries, the results of which, ultimately, are aimed at improving the quality of life of the population (autonomy of mobile devices, the development of information and communication technologies, electric transport, etc.).

As part of a research project "Technological Forecasting using Data Mining and Semantics" (MIT, USA) methods based on semantic technologies are being developed to analyze an array of scientific and technical information from electronic resources, including in the field of advanced technologies of renewable energy sources and distributed energy generation [4]. Work is also being done on forecasting the development of the application of individual technologies in the energy sector [5].

Successes have been achieved in the creation of intelligent search systems, technologies for reconciling heterogeneous information, automatic generation of domain taxonomy, methods for identifying promising technologies and identifying their innovative indicators (quantitative and qualitative), methods for clustering and visualizing cognitive research and development maps [6-8]. It is worth noting that intelligent approaches in the field of forecasting innovative development are still in their infancy and have problem points: quality and reliability of information sources, inconsistency of data (noise, irregularity), and problems of integration of heterogeneous data. It is necessary to improve the mechanisms of expert or other verification of the quality of the results. These approaches can produce although quite superficial, but interesting solutions.

III. ENERGY TECHNOLOGY FORECASTING

In Russia the tools used to support decision-making on scientific and technological forecasting and innovative development in the energy sector can be roughly divided into two main groups of approaches. The first is based on the methods of the scientific and technological foresight, when all reviews, assessments, forecasts and recommendations are formed on the basis of the consolidated opinion of experts [9]. Despite the simplicity, efficiency, low costs, this approach is often criticized for subjectivity, weak validity and lack of responsibility of experts. The core of the second group of methods is the system analysis methodology, which is the most strictly scientifically based approach, since it uses as the main tools: (a) methods of system assessment and technology comparison; (b) system technological modelling (mathematical models of technologies and their structures). At the same time, energy technologies themselves are taken into account in the totality of their basic characteristics, as well as a whole range of external factors (needs, resources, ecology, functioning, institutions, etc.) [10]. However, the use of such tools for the development of innovative development programs, including in the energy sector, implies the implementation of a complex of full-scale research work that requires a lot of labour and time of highly qualified research teams. Despite the fact that the results of such scientific research can be of very high quality and give excellent solutions for their specific industry scientific and technical problems, their use is limited. The main problems are the lack of effective mechanisms for the operational adaptation of the system technological modelling apparatus to changing conditions of technological development, especially in terms of taking into account factors of scientific and technological progress in related or loosely coupled areas, in the complexity of taking into account interdisciplinary and infrastructural technological effects, in the increased information, technical and organizational complexity objects of study.

On the other hand, in recent years Melentiev Energy Systems Institute of Siberian Branch of the Russian Academy of Sciences (ESI SB RAS) has successfully developed intelligent methods and tools to support decision-making in fairly close areas of energy research, namely, in researching the development directions of the country's fuel and energy complex taking into account energy security requirements [11]. Most of the methods used are based on intelligent semantic technologies for searching, extracting and analyzing heterogeneous data from electronic information sources in line with the Big Data concept [12, 13]. In particular, the Tech Mining methodology is proposed and applied in practice as a form of statistical contextual analysis of text documents based on scientific and technical resources (primarily scientific publications and patent databases) to identify breakthrough research and development and to characterize their innovative potential [14]. The Tech Mining technique is based on an approach called Text Mining, which is used to algorithmically identify and detect correlations, relationships, and previously unknown practically useful interpreted knowledge in text data fragments [15-17], which can be used to make decisions in various fields.

As a result of the author's analysis of scientific publications on advanced studies of the energy field was revealed that, against the backdrop of the world background, China has sharply increased the number of scientific works in the field of engineering, energy and materials science over the past 5 years. This indicates the scientific research and engineering work carried out by our eastern colleagues to improve the existing creation of new mobile sources of electricity (high-capacity batteries) and, possibly, superconductors. The implementation of industrial designs will allow China to confidently enter, for example, the electric car market place. The distribution of works in the field of Text Mining (more than 11,000 works in 4 years) and Tech Mining (400 works in 4 years) is also quite interesting. The largest contribution to Text Mining is made by American scientists and is mainly focused on the development of these methods in relation to social media [18, 19]. And Tech Mining is more developed by Chinese scientists for the tasks of scientific and technological forecasting [20-22].

IV. ENERGY TECHNOLOGY FORECASTING TOOLS

Intelligent Information Environment of Energy Innovative Development Forecasting (IIEIDF) was developed to support the conduct of scientific research in the field of forecasting the development of energy at the ESI SB RAS [23-25]. This information system is able to accumulate process, classify information, and store descriptions in the warehouse and issue analytical reports. In addition, the system implements an environment to support the work of scientific groups. As noted above, Open Data and Big Data are used as sources of information for making forecasts of scientific and technological development.

Information is collected both from structured databases of scientific publications and developments, as well as from ordinary Internet resources. A generalized scheme for processing information is shown in figure 1. At the first stage, text data is extracted from the documents to be analyzed. Then the keywords and word-sets are highlighted, which should be consistent with the concepts from the ontology. At the last stage, based on the agreed data and their numerical characteristics, document digests are worth. IIEIDF uses ontologies to create a single conceptual basis for integrating both expert knowledge and for integrating the results of scientific and technical activities.

The main components of IIEIDF and their interaction are presented in figure 2. Many data adapters are used to extract information from structured information sources. The main task of each adapter is to obtain information from a structured source and prepare it for analysis by converting the general format. Many crawlers are used to search for information on the Internet. Crawlers are combined

into farms. Each crawler enters a web page, analyzes it, extracts text and links to other pages or documents. Next, the crawler goes to analyze the extracted links. We need the initial set of links to resources to start the crawler farm, which can be collected in automatic or manual modes. Links are extracted from infobases of well-known Internet search engines automatically. 396 initial keywords and phrases were used to study renewable energy and new materials in the energy sector. A database of scientific publications and

developments from more than 450 thousand documents was collected to inform the scientific project. The main task of processing documents after their search is from classification and integration into a single warehouse. Integration of document descriptions is based on the ontology system. Further access to information in IIEIDF is done through ontologies. New keywords and expressions can be detected by analyzing the document, which must be added to the ontology to enrich the knowledge base.
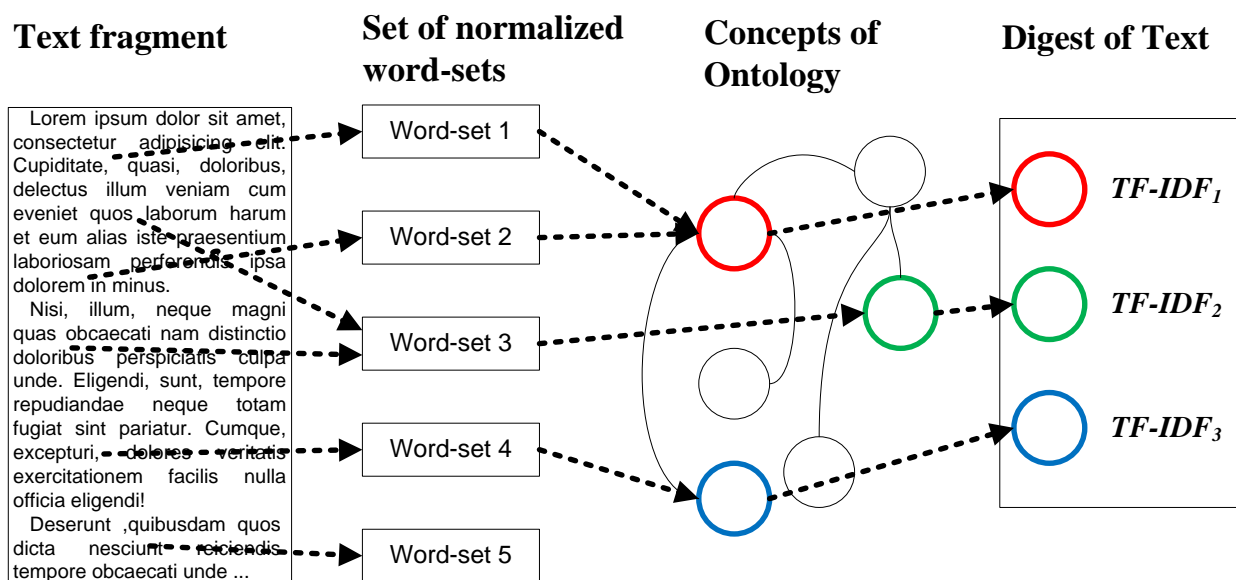

Fig. 1. The generalized scheme of processing and classification of information.
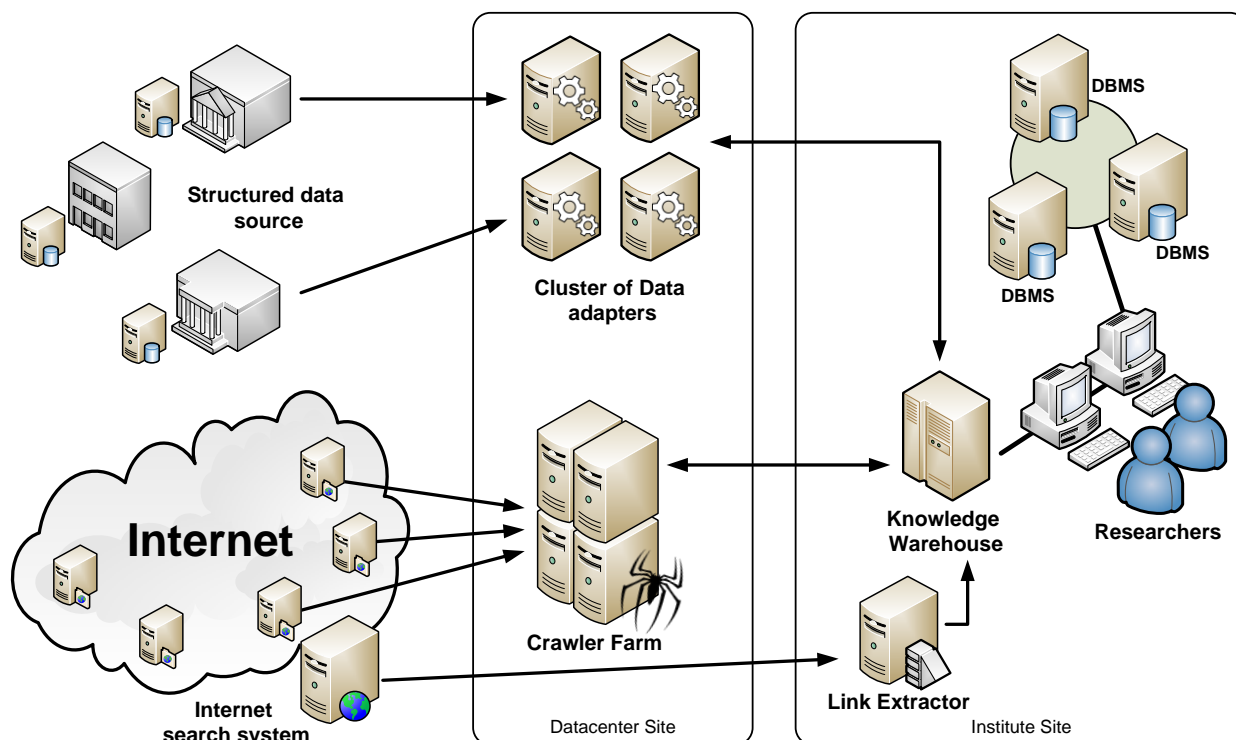

Fig. 2. Architecture of Intelligent Information Environment of Energy Innovative Development Forecasting.

The most valuable services are located in ESI SB RAS. Data adapters and crawlers required fast Internet connection and low latency. Therefore, these components were being placed in Datacenter. Data exchange between sites is not large by reason of applying local copy of Ontology database. One important consequence of the selected architecture is an easy increasing count work-servers. The cost of renting Virtual Dedicated Servers (VDS or VPS) in datacenters is not so expensive. Usually virtual servers are paid monthly and the price ranges from $2 to $10-$15 per one per month. Thus, computing resources for collecting information can be quickly and fairly cheaply increased by leasing virtual servers if necessary. As shown by the experiments, the installation and launch time of the virtual server takes no more than 4 hours from the payment.

V. TECHNOLOGY OF ENERGY INNOVATIVE DEVELOPMENT FORECASTING

To improve the efficiency and quality the forecast estimation author was developed technology of application of IIEIDF. This technology consists of four main parts and several tasks within:

1. *Preparation for the study.* At this stage, the initial environment configuration is performed. Researched are attracted as experts to create the configuration of the future system. IT professionals, knowledge engineers and linguists should identify key features of IIEIDF configuration.

2. *Data collection.* The infrastructure is being deployed and information is being collected by crawlers. IT professionals should deploy selected solution and start the process of data collection.

3. *Data analysis.* Analysis is performed based on the collected and processed data. Experts should use the accumulated and integrated knowledge from the warehouse to update their models and conduct qualitative and quantitative research.

4. *Process of results analysis.* It is the final stage at which researchers process the results of experiments. Experts should use tools to process the results of experiments, identify trends, design research results, as well as their integration into knowledge warehouse.

Below we take a closer look at each individual stage of technology.

On the first stage research team should make some activities to prepare for the study:

1. *Designing of ontology system components.* Knowledge engineers together with experts should develop a system of ontologies; define basic concepts and links between them.

2. *Constructing of initial terminological reference list (manually or machine learning).* Linguists together with experts should define basic linguistic concepts and keywords of the domain

3. *Binding word-sets with ontology concepts (elements of the ontology system).* At this stage, knowledge engineers should map the resulting term lists with intermediate notions of the ontology system.

4. *Identification of information sources.* Domain experts should determine the basic sources of structured and semi-structured information that will be used to populate the warehouse.

5. *Designing of data adapters and preparation of data flow.* IT professionals should develop data adapters to extract key information on each specific type of object from sources.

6. *Planning of service infrastructure for data collection.* IT professionals should develop a scheme of service infrastructure; identify the deployment points of the main compute nodes, crawler farms and adapter clusters.

The second stage performed by IT professionals and contains activities related to data collection:

1. *Deploying network service infrastructure.* Server installation and configuration farms of crawlers and data adapters.

2. *Collecting of initial points of data retrieval of free format sources.* For this, we use a list of requests is used and well-known search engines for getting a list of starting URL-addresses in automatic mode.

3. *Collecting documents with using automatic classification methods* (automatically using IIEIDF components). At this stage crawlers and data adapters are used, which perform a preliminary analysis of documents from the Internet.

4. *Identification of new keywords to enrich the ontology system* (automatically using IIEIDF components).New keywords and word expressions can be found after analysis based on the frequency of occurrence with already known elements.

5. *Clustering of documents and building thematic collections* (automatically using IIEIDF components). Documents are characterized by created digests, which briefly describe their internal semantic content. These digests are used to divide the entire set of documents into categories.

6. *Organizing of access to knowledge via ontology system* (using IIEIDF components). Ontology elements can be used to access digests and transitively to original documents.

The third stage contains activities of data analysis:

1. *Validation of exists mathematical and technical models.* Researchers using IIEIDF components perform a search for related knowledge based on the concepts of ontology systems.

2. *Performing computer experiments.* Researchers using updated models conduct qualitative and quantitative research

3. *Storing results in data warehouse.* Researchers describe the obtained results with the concepts of the ontology system and save the results to the knowledge warehouse.

4. *Creating graphical models and representing result in visual form.* Researchers use visual analyst tools to build models.

The last stage related to process of results analysis:

1. *Analysing existing trends and result of computer experiments.* Chiefs of expert groups analyse trends and research results; formulate a set of hypotheses regarding possible situations.

2. *Investigating and testing of hypothesis.* Researchers test hypotheses using quality analysis tools, modelling systems, match them with the knowledge in the warehouse

3. *Preparing science report with using collective tools.* Researchers draw up the results of the research in the form of complete documents containing formal sections, which are described by the concepts of the ontology system. The system of collective work on scientific reports can be used for these purposes.

4. *Registration of research results within knowledge warehouse.* Created scientific reports and their descriptions may be placed in the warehouse and compared with concepts from the ontology system, which in the future will make it easy to find these results.

The steps of the technology presented should be performed sequentially. It is possible to perform many researches on the collected data after a single execution of the first and second parts. A warehouse of 450 thousand documents was collected for research on the development of energy technologies based on this technology.

VI. CONCLUSION

The use of the information system and technology for the applying of tools presented in the work allows reducing the time for developing analytical reports on promising research in the field of energy. The main stages of the technology include creating an initial terminology list and developing an ontology system, searching for information sources and developing adapters. The proposed technology was tested during research in the field of energy and allowed to identify the main trends of research.

REFERENCES

[1] V. Coates, et al. On the future of technological forecasting // Technol. Forecast. Soc. Change, 67 (1), 2001. – pp. 1 – 17.

[2] Zhang, Y., et al. Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research // Technol. Forecast. Soc. Change, 2016. – DOI: 10.1016/j.techfore.2016.01.015.

[3] Future-Oriented Technology Analysis. Strategic Intelligence for an Innovative Economy / Eds. C. Cagnin et al. - Springer, 2008, 170 p. – DOI: 10.1007/978-3-540-68811-2.

[4] Woon, W.L., Zeineldin, H., Madnick, S. Bibliometric analysis of distributed generation // Technol. Forecast. Soc. Change, vol. 78(3), 2011, pp. 408–420. – DOI: 10.1016/j.techfore.2010.08.009

[5] Chanchetti L. F., Diaz S. M. O., Leiva D. R. et al. Technological Forecasting of Hydrogen Storage Materials using Patent Indicators / International Journal of Hydrogen Energy, vol. 41, Elsevier, 2016, pp. 18301-18310. – DOI:10.1016/j.ijhydene.2016.08. 137.

[6] Woon W.L., Aung Z., Madnick S. Forecasting and Visualization of Renewable Energy Technologies using Keyword Taxonomies. // Data Analytics for Renewable Energy Integration. / Ed. by W.L. Woon et al. Second ECML PKDD Workshop (DARE 2014), Nancy, France, September 19, 2014. Revised Selected Papers, Springer, 2014, pp. 122–136. – DOI: 10.1007/978-3-319-13290-7_10.

[7] Nizhegorodtsev R. M., Ratner S. V., Trends in the Development of Industrially Assimilated Renewable Energy: The Problem of Resource Restrictions, Thermal Engineering, Springer, 2016, pp. 197-207. – DOI:10.1134/S0040601516030083

[8] Daim T.U., Chiavetta D., Porter A.L., Saritas O. Anticipating future innovation pathways through large data analysis / Innovation, technology, and knowledge management, Springer, 2016, 360 p. – DOI 10.1007/978-3-319-39056-7.

[9] Forecast of scientific and technological development of Russia: 2030 / Ed. L.M. Gokhberg, S.P. Filippova. – Moscow: Ministry of Education and Science of the Russian Federation, National Research University Higher School of Economics, 2014, 244 p.

[10] Systems research in the energy sector: retrospective of scientific directions SEI-ESI / Chif ed. N.I.Voropai. – Novosibirsk: Nauka, 2010, 686 p.

[11] Massel L.V., Massel A.G. Intelligent computing in studies of energy development trends // Control, computer engineering and informatics / Bulletin of the Tomsk Polytechnic University,.2012, vol. 321, No 5, pp. 135-141.

[12] Mayer-Shenberger V., Cukier K. Big Data. A Revolution That Will Transform How We Live, Work and Think. Eamon Dolan/Mariner Books; Reprint edition, 2014, 272 p. ISBN-13: 978-0544227750

[13] Marz N., Warren A.J. Big Data: Principles and Best Practices of Scalable Real-Time Data Systems. Wiley, 2015, 360 p. – ISBN: 978-1118961742

[14] Cunningham, S.W., Porter, A.L., and Newman, N.C. Tech Mining, special issue of Technological Forecasting & Social Change, 73 (8), 2006, pp. 915- 1060.

[15] Mirhosseini M. A Clustering Approach using a Combination of Gravitational Search Algorithm and k-Harmonic Means and its Application in Text Document Clustering / Turkish Journal of Electrical Engineering and Computer Sciences, vol. 25, no. 2, 2017, pp. 1251-1262. DOI:10.3906/elk-1508-31.

[16] Zheng L., Noroozi V., Yu P. S. Joint Deep Modeling of Users and Items using Reviews for Recommendation, 2017. DOI:10.1145/3018661.3018665.

[17] Zhang M. L, Zhou Z.H. Multilabel Neural Networks with Applications to Functional Genomics and Text Categorization / IEEE Transactions on Knowledge and Data Engineering,

vol. 18, No. 10, 2006, pp. 1338-1351. – DOI: 10.1109/TKDE.2006.162

[18] Ragini, J. R., P. M. R. Anand, V. Bhaskar. Big Data Analytics for Disaster Response and Recovery through Sentiment Analysis / International Journal of Information Management, vol. 42, 2018, pp. 13-24. – DOI:10.1016/j.ijinfomgt.2018.05.004.

[19] Pereira, R. B., et al. Categorizing Feature Selection Methods for Multi-Label Classification / Artificial Intelligence Review, vol. 49, no. 1, 2018, pp. 57-78. DOI:10.1007/s10462-016-9516-4.

[20] Huang, Y., et al. A Hybrid Method to Trace Technology Evolution Pathways: A Case Study of 3D Printing / Scientometrics, vol. 111, No. 1, 2017, pp. 185-204. DOI:10.1007/s11192-017-2271-8.

[21] Peng, H., et al. Forecasting Potential Sensor Applications of Triboelectric Nanogenerators through Tech Mining / Nano Energy, vol. 35, 2017, pp. 358-369. – DOI:10.1016/j.nanoen.2017.04.006.

[22] Zhang, Y., et al. Technology Roadmapping for Competitive Technical Intelligence / Technological Forecasting and Social Change, vol. 110, 2016, pp. 175-186. – DOI:10.1016/j.techfore.2015.11.029.

[23] Mikheev A.V. Ontology-based Data Access for Energy Technology Forecasting / Advances in Intelligent Systems Research, Vol. 158, 2018, pp. 147–151. – DOI: 10.2991/iwci-18.2018.26.

[24] Kopaygorodsky A.N. Methods and technologies of building an intelligent service for energy technology forecasting / International Scientific Journal Industry 4.0, 2018, Issue 5. Publisher: Scientific Technical Union of Mechanical Engineering "Industry 4.0". pp. 223-228.

[25] A. Kopaygorodsky. Ontology-Based Decision Support System for Forecasting of Energy Infrastructure Development // 3rd Russian-Pacific Conference on Computer Technology and Applications (RPC), Vladivostok, 2018, pp. 1–5. DOI: 10.1109/RPC.2018.8482172