Information Support of Modeling the Functioning of Defense Industries

A.A. Akinshin[0000-0001-6487-5760] and Yu.E. Polak[0000-0001-8411-335X]

Federal State Budgetary Institution of Science Central Economics and Mathematics Institute of the Russian Academy of Sciences, 47 Nakhimovski Pr., Moscow 117418 Russia
aaa@cemi.rssi.ru, polak@cemi.rssi.ru

Abstract. When processing huge amounts of information arising in the process of activities of individual organizations and entire sectors of the economy, new approaches and solutions are needed. For effective conversion of national defense industry, an information and analytical system is being developed. The paper discusses the created toolkit for systematization of open data on military products and defense industry potential. In particular, a resource is being developed that will be a kind of aggregator of this information.

Keywords: Military-industrial Complex, Scientific and Industrial Policy, Information Support, Network Information Resources, Databases.

1 Introduction

At present, the military-industrial complex of Russia includes about 800 industrial enterprises and about 570 design, engineering and technological organizations. About 1200 enterprises and organizations (including repair plants of the Russian Ministry for Defense) are located in 70 Russian regions. The number of people employed in enterprises and organizations of the defense industry exceeds 2 million.

The defense industry, including military products, accounts for more than 70% of all scientific production in Russia, it employs more than 50% of all research workers. The military-industrial complex is the most high-tech sector of the country's economy [1].

The modern defense industry complex as a set of specific economic entities is a complex system with a large degree of diversity and uncertainty, with complex management that defines many options when choosing decisions on functioning. Accordingly, the problem of increasing the efficiency of development management of the defense industry sectors is also complex. In order to analyze, model and predict this process in modern conditions, taking into account the growth of production volumes and innovative modernization, it is necessary to develop a comprehensive, scientifically sound and practically applicable methodological apparatus and tools.

2 Project Features

The Department for Economic Informatics of the Central Economics and Mathematics Institute of the Russian Academy of Sciences is entrusted with the implementation of the project 'Information-analytical and software tools to systematize available data on military products, determine and analyze the defense industry potential for the innovative growth of the national economy'. This study should assess the current state of the Russian defense industry and its role in the economic development of the country, identify structural problems and threats to its effective development, and assess possible trends in the further development of domestic enterprises cooperating with this sector of the economy.

A feature of the developed in the project information-analytical and software tools, as well as modeling technologies is the use of multi-level decomposition of the national innovation system and the process of its modeling. The number of levels is determined by the complexity of the simulated object and the degree of detail of the constructed model. The approach used will allow to present the national innovation system in the form of an arbitrary decomposition set of subsystems and elements, and to divide the process of its functioning into its constituent procedures. The conceptual modeling tools created within the framework of the project will help to improve the quality of the designed and functioning defense industries, reduce the overall costs and terms of their reform, and increase the productivity of designers and managers.

As mentioned above, the military-industrial complex is one of the most modern and high-tech sectors of the domestic industry, an innovative driver of the economy. For the implementation of an effective scientific and industrial policy in our country there are necessary prerequisites. This is the presence of advanced science, a developed education system, production base, human potential, financial and material resources, etc. However, in the conditions of the transition period and the protracted crisis of the real sector of the economy, new challenges arise. In particular, the scientific, technical and production potential of the defense industry is adversely affected by low energy supply, depreciation of fixed assets, low utilization of production capacities, the collapse of existing cooperative ties, reduction in the number and quality of personnel, rising energy prices and so on. Researchers are faced with the task of giving a comprehensive assessment of the current state and role of the defense industry in the economic development of the country; identification of structural problems and external threats to its effective development.

For this, it is necessary to create information, analytical and software tools to systematize the available open data on military products. Important from a methodological point of view, the areas of research are the allocation and description of the studied subject area; preparation of clear methods and recommendations to ensure the implementation of theoretical principles in practice; the creation of information and software tools that allow to automatically design and analyze conceptual models of various innovative objects and their complexes.

3 Project Information

To build classifiers and taxonomies, for the subsequent filling and updating of the information-analytical system, the following most important sources of information are considered:

• electronic collections and libraries;
• monographs, dissertations;
• materials of conferences and seminars;
• scientific and technical periodicals;
• individual publications, articles;
• reports;
• patents;
• websites of various levels (personal pages, official websites of enterprises, industrial groups,
clusters etc);
• news feeds;
• lists of resources, classifiers.

Open sources are structured in terms of their relevance, completeness, relevance and regularity of updates. Thus, there is a need to rank them for all these parameters.

The team of performers has the necessary experience in developing significant information projects. Since 1995, the Laboratory for Network Information Resources of the CEMI RAS has been studying the information resources of the domestic Internet. In a short time, it became one of the leading collectives of the country in this area. Its main product is the database ‘Internet in Russia, Russia on the Internet’ (1996), which became the basis for ‘Au!’, famous online directory of Internet resources [2]. In 1997, the database was registered in RosAPO, and in 2000 – in the Scientific and Information Center ‘Informregister’. The laboratory was preparing analytical, educational and methodological materials about network information resources for professionals; developing online annotated indexes of web addresses, as well as printed directories. For this purpose, an analysis of rubricating systems of electronic documents and directory structures of electronic libraries was made. The possibilities of facet classification, taxonomy and metadata systems, the development of navigation systems for scientific and technical information resources, the role and importance of the information catalog in the architecture of the data warehouse, various architectural solutions for constructing metadata management systems were investigated.

This experience finds application in the development of information-analytical and software tools to systematize open data on military products and the defense industry potential. We emphasize: focusing exclusively on open sources is our principled position. The development team is composed of first-class professionals, who at one time held senior positions in the Soviet and Russian army. Their qualifications and experience in the implementation of previous projects give confidence that scientists of the RAS are quite capable of making sound strategic forecasts and plans for the development of defense industries based on available open data.

Information analytical system should include

• knowledge base;

• regulations;
• statistical information;
• examples and technical specifications of successful developments;
• help desk on Internet resources.

Performing research and analytical work, specialists are interested in having a universal information resource containing all of the above. It is very convenient if all these materials are collected in one place and have an intuitive and uniform interface. And in order not to overload the portal with second-rate information, it should be equipped with a catalog of links – a navigator for open data sources. It should be noted that the official websites of defense organizations do not usually contain such complex information. While doing webometric research, one of the authors made sure [3,4] that the websites of higher education institutions of law enforcement agencies are very concise in comparison with civilian universities (however, the reasons for this are understandable). An example here is the US Department of Defense portal Defense.gov.

The system must perform such functions (operations) as

• search for objects by metadata elements and their combinations with a relatively developed query language,
• navigation between interconnected objects,
• the formation of sets of objects with specified characteristics.

An analysis of information sources will help determine the main reasons for the lack of competitiveness of Russian high-tech enterprises and highlight the factors for its increase.

4 Project Implementation Procedure

Among the main functional tasks of information and analytical tools should include a meaningful analysis of the problem situation, determining the parameters of the management strategy through the procedures of refined development programs, as well as accumulating experience in automation and application of technology in practice. The need for constant adaptability of planning technology to the course of economic transformations and information flows puts forward stringent requirements for the adaptive properties of tools and the corresponding intellectually rich knowledge base. By influencing the control parameters of decision-making procedure, the decision-maker receives various options for program-planning decisions that differ in technical and economic indicators, the degree of use of local and system-wide resources, the degree of fulfillment of orders for products (including scientific), etc. Each of these design decisions is rational in terms of a specific criterion. A set of models of information and analytical technology, as well as methods for their practical implementation, should ensure the efficiency of calculations, both on aggregated and detailed source data.

The adoption of decisions aimed at optimizing the scientific and industrial policy in general and the military-industrial complex in particular should be based on an adequate information and technological base.

Consider the main stages of the project. Decomposition of subject area description. Decomposition of modeling processes. Building a taxonomy. For example, a two-stage semi-automatic method. At the first stage - building the basis of taxonomy, two or three upper levels in accordance with regulatory documents formalizing the subject area under consideration. At the second stage - step-by-step completion of taxonomy topics based on analysis of open source data, using a measure of similarity (relevance) to that text.

To build an information-analytical system it is necessary:

1. Develop a knowledge base model that allows to collect, store, process data on products of the military-industrial complex obtained from open sources.

The knowledge base is built on a database. One of the main goals is the technology base. When it is created, several classifications (taxonomies) arise when developing a database model.

‘Cluster-industry’ approach. It is proposed to use the Rostec classification, and then the first levels look like this: ‘OPK’ → ‘Clusters’ → ‘Industry Associations’ → ‘Enterprises’ → ‘Products (type / group)’ → ‘Particular Product’ → ‘Components’. It is relevant to supplement Rostec clusters (aviation, radio electronics, and weapons) so that they cover the widest possible range of defense industry products.

‘Technological’ approach. Classification based on technologies used in the production of defense industry products. The taxonomy is based on the international patent classification.

Application area. The base should allow both the initial creation of appropriate taxonomy, and its further expansion in the process of filling and operation. Then, when constructing the ER-model of the base, three sets of attributes arise, corresponding to the above.

The decomposition of the subject area (‘defense industry products’) should be carried out according to different types of criteria at different levels. Criteria for ‘industry’ classification should be used from the level of ‘defense industry’ to the level of ‘Product’; from the level of ‘Product’ - technological classification and scope.

2. Prepare a methodology for the collection, systematization and analysis of data on the products of the military-industrial complex, as well as various options for its presentation, depending on the tasks being solved.

3. To develop software tools those allow the collection and analysis of data, their subsequent presentation.

5 Some Relatively New Methods

To solve some of these problems, the method of cognitive modeling will be used. The basic principles of the cognitive approach in the decision-making process were first formulated by R. Axelrod [5], at the same time the initial version of the mathematical apparatus for modeling was proposed. The methodological basis for creating conceptual models in the study of dynamic systems in economics was laid by J. Forrester [6, 7]. Cognitive methodology has also been successfully applied by domestic researchers.

The choice of this method as one of the most important for this study is due to a number of important points. The proposed approach is focused on the qualitative analysis of complex situations, interpreted as weakly structured systems, characterized by a lack of accurate quantitative information about the processes occurring in them, as well as including various ‘qualitative’ variables. The number of variables in such situations can be measured in dozens, and all of them are woven into a web of causes and effects. It is extremely difficult to see and grasp the logic of the development of events in such a multifactor field. At the same time it is necessary to make decisions continuously on the choice of certain measures contributing to the development of situation in the right direction. The main advantage of the proposed methodological apparatus is the possibility of systematically qualitative account the remote consequences of decisions made and identifying side effects that may prevent the implementation of seemingly obvious solutions and which are difficult to evaluate intuitively with a large number of factors and a variety of interaction between them.

Processing of huge amounts of information arising in the process of activities of individual organizations and entire sectors of the economy, in particular, the defense industry, requires new approaches and solutions. Here it is appropriate to use the technologies and methods of Big Data. This term, unlike many others, has a specific date of appearance - September 3, 2008, when the paper [8] had appeared in Nature. C.Lynch then proposed for the new paradigm the special name Big Data, to reflect the transition of quantity into quality.

This problem existed long before its reflection in the mass media: more than 60 years ago it had arisen in space research, in problems of nuclear physics. Big data appears

- in the presence of large volume, which is problematic for processing facilities;
- at the demand of a high processing speed, which the processing facilities cannot provide;
- with high complexity and diversity that doesn’t allow to use traditional processing tools.

The processed data can be used both for analysis and for forecasting, while the areas of application of Big Data solutions are so wide that they penetrate deeper into everyday life. Continuous data growth has now affected wide areas of activity and will increase exponentially. Therefore, it is not possible to organize the process associated with their receipt, management and processing for the required time with traditional tools.

Working with Big Data is not cheap, but the costs will be meaningless without clear understanding of desirable result. In addition, for their analysis, a specialist who can customize the machine model and regularly reconfigure it is needed.

Programs focused on processing large amounts of data deal with data files in tera- and petabytes. In practice, this data comes in a variety of formats and is often shared among several storage sources. Processing of such data sets usually occurs in a phased analytical pipeline mode, which includes the stages of data conversion and integration.

Giant information volumes in combination with high speed require the appropriate computers, so almost all major manufacturers offer specialized software and hardware: SAP HANA, Oracle Exadata Database Machine, Teradata Extreme Performance Appliance, NetApp E-Series Storage Technology, IBM Netezza Data Appliance, EMC Greenplum , HP Converged Infrastructure.

The analysis of large volumes of data requires the involvement of technologies and means of implementing high-performance computing. Artificial intelligence methods are applied, including machine learning, information extraction, and data mining. The main problems are the complexity and physical volume of the information collection. It should be noted that the actual data processing also includes the construction of the algorithm and the time for its description and debugging. Unique data collections require the development of unique algorithms, which dramatically increases the total processing time.

It is often necessary to solve numerous problems arising when working with unstructured information from various sources of social networks, blogs, forums, news sites, etc. Text analytic methods (morphological, syntactic, semantic) provide the ability to control this stream and extract the maximum possible benefit from it.

Extracting information from unstructured text allows performing text data mining. Text mining covers new methods for performing semantic text analysis, information retrieval and control. A synonym for text mining is KDT (knowledge discovering in text). Like most cognitive technologies, text mining is an algorithmic identification of relationships and correlations in existing textual data. Data mining is a process of non-trivial extraction from data of previously unknown and potentially useful information [9]. An important task of the text mining technology is extraction from the text its characteristic elements, which can be used as document metadata, keywords, annotations.

Among standard text mining applications are such mechanisms of ‘classical’ data mining as automatic text classification with the ability to automatically filter certain terms; algorithms and methods of clustering, factor analysis; automatic search in documents for given words or phrases.

Data mining methods can be applied to data collected from experiments in many scientific fields (astronomy, physics, medical imaging, bioinformatics). In the tasks of systematization of data on military products, analysis of the military-industrial potential, this method has not been previously used.

6 Conclusion

The implementation of the above will allow to approach the solution of the fundamental problem of forming an effective scientific and industrial policy and implementing its tools in the context of structural transformations in the modern innovative and high-tech sector of domestic industry and, based on the example of the military-industrial complex, scientifically substantiate a specific system of measures that provide progressive structural and technological shifts in complex and in the Russian economy as a whole. An ER database model is being developed; the methodology for collecting and systematizing information is being tested.

The article uses the materials of the report at the XXI All-Russian Scientific Conference ‘Scientific Service on the Internet’ (September 2019).

Acknowledgments. This work was financially supported by the Russian Foundation for Basic Research in the framework of the scientific project Nos. 18-00-00177 (18-00-00172)

References

1. Hrustalyov, E.Yu.: Oboronno-promyshlennyj kompleks Rossii: prednaznachenie, sostoyanie i perspektivy razvitiya. Nacional'nye interesy: prioritety i bezopasnost'. 7(35), pp. 61–71 (2011).
2. Kulikov, V.V., Polyak, Yu.E.: Katalog russkoyazychnykh resursov interneta ‘Au!’. In: Telematika '98. Sbornik nauchnykh trudov, pp. 334–335. LITMO, SPb (1998).
3. Polyak, Yu.E.: Rossiyskiy i mezhdunarodnyy opyt vebometricheskikh issledovaniy. Informatsionnyye resursy Rossii. 6 (142), pp. 2–9 (2014).
4. Polyak, Yu.E.: Otsenivaniye i ranzhirovaniye veb-saytov. Vebometricheskiye reytingi. Nauchnyy redaktor i izdatel. 2(1), 19–29 (2017).
5. Structure of Decision. The cognitive Maps of Political Elites. R. Axelrod (ed.), pp.1–422. Princeton University Press, N.Y. (1976).
6. Forrester, J. W.: Industrial dynamics – after the first decade. Management Science, 14(7), pp. 398–415 (1968).
7. Forrester, J. W.: Urban Dynamics. Pegasus Communications: Waltham, MA (1969).
8. Lynch, C.: How do your data grow? Nature 455(7209), pp. 28–29 (2008).
9. Olmer, P.: Knowledge discovery: data and text mining. In: Inverted CERN School of Computing, pp. 47–55. CERN, Geneva (2006).