The following results were achieved in 2018 in RSF project #18-11-00302.

1. **The concept of creation and functioning of the intellectual systems of analytical processing of digital network content (ISAPDNC) for the purpose of detection and counteraction of inappropriate, dubious and harmful information** is developed. The concept includes: main definitions; the principles of the organization of intellectual analytical processing of the digital network content extended in the Internet and on social networks; purposes and problems of ISAPDNC; the features of Internet space and social networks influencing creation of ISAPDNC; mechanisms of ensuring intellectual analytical processing of digital network content on the Internet and on social networks; structure of ISAPDNC; requirements to ISAPDNC. The main objective of functioning of ISAPDNC determined identification in information objects (IO) of the Internet and on social networks of inappropriate, dubious and harmful information and development of measures of counteraction to these types of information. The principal ISAPDNC structural components are defined: (1) the distributed intelligent scanners and components of preprocessing; (2) the components of multidimensional assessment and a categorization of IO including basic qualifiers, aspect qualifiers and final qualifiers; (3) components of ensuring timeliness of the analysis of IO; (4) components of elimination of incompleteness and inconsistency of results of assessment and categorization; (5) components of adaptation and retraining of a system; (6) the components of development and choice of measures of counteraction of undesirable, doubtful and harmful information including base of counter-measures and the module of support of decision-making; (7) components of implementation of visual interfaces. For each ISAOTSSK structural component the carried-out functions are defined and requirements imposed to them are proved.

2. **The general approach and requirements imposed to components of collecting and preprocessing of network information objects on the basis of use of the distributed intelligent scanners** are developed. An approach is based on decentralized control by a complex of the distributed intelligent scanners which are independently organized for execution of separate operations of computation process. Initial data for collecting and preprocessing of network IO are defined: (1) a finite set of the network addresses for scanning; (2) the unique network addresses of arrangement of network IO for their loading and preprocessing; (3) the list of keywords for search of network IO. The following main properties of network IO are selected: targeting, variability, availability, structurability and hierarchy. Mechanisms of ensuring availability of network IO depending on their level are offered. Functional requirements to components of collecting and preprocessing of network IO are proved: (1) detection and loading of network IO taking into account parameters of the set segment; (2) typization of lexemes of the semi-structured and structured network information content; (3) structural categorization of IO; (4) identification and overcoming locking gears of the automated collection of information; (5) prioritizing of IO. Nonfunctional requirements are: flexible scaling, decentralized control, maximum efficiency; high resource intensity.

3. **The general approach and requirements imposed to components of multidimensional assessment and a categorization of semantic filling of information objects** are developed. Approach is oriented to a solution of a problem of automatic classification of network IO with use of the data collected in distributed DB (DDB) of IO by means of intelligent scanners. Stages of a solution of this task are: preprocessing of the text; extraction of signs; reduction of dimension of space of signs; training of the qualifier by methods of machine learning; qualifier

assessment. Preprocessing of the text includes tokenization and lemmatization, removal of a stop word, stemming and morphological analysis. At a stage of extraction of signs the numerical model (indexation) of the text is formed. Reduction of signs is reached due to removal of strongly correlating and not informative values of attributes. For assessment of the text qualifier standard metrics of accuracy are used. The main metrics are completeness metrics, however metrics of accuracy and F-measure are also considered. Functional requirements define need: the support of preprocessing of text maintenance of IO; the support of the analysis of structure of a HTML page; the support of the analysis of digital contents; the support of processing of additional information on IO; identification of technologies on server side of the corresponding IO; extraction of the executed scenarios. A set of selected nonfunctional requirements consists of: accuracy, flexibility (adaptivity), high-speed performance and scalability.

4. **The general approach and requirements imposed to components of ensuring timeliness of the multilevel and multi-modular analysis of IOs on the basis of use of parallel computing** are developed. An approach is oriented to parallel processing of data in the ISAPDNC components. Taking into account advantages and shortcomings of different ways of parallelization of calculations, the decision to use in the project for ensuring timeliness of the analysis of IO by means of machine learning the Spark Hadoop platform and different algorithms from its Spark MLlib library is made. The main requirements to components aimed at providing operability of all system which consist in ensuring high scalability and fault tolerance and also the organization of the distributed data storage are proved.

5. **The general approach and requirements imposed to components of elimination of incompleteness and inconsistency of assessment and a categorization of semantic filling of IOs on the basis of use of methods of processing of incomplete, contradictory and indistinct knowledge** are developed. An approach comes down to improvement of qualifiers of IOs with use methods of probabilistic, indistinct and neural network estimation of situations, fuzzy and neural network inference and fuzzy and neural network optimization. The structure and levels of architecture of components of elimination of incompleteness and inconsistency taking into account decision-making "chain" in the conditions of uncertainty is offered. Decision-making at each level is carried out in the conditions of the uncertainty caused by not stationarity of functioning of ISAPDNC, influences of the destabilizing factors, the external environment, illegibility of the purposes and inconsistency of problems of assessment and categorization and also other factors. Private and generalized settings of tasks on development of models and methods of elimination of incompleteness and inconsistency of assessment and categorization of digital content are formulated. A set of their number consists of: (1) the probabilistic methods allowing to eliminate aspects of a class of stochastic uncertainty; (2) the methods of the theory of fuzzy sets allowing to eliminate ambiguity, uncertainty, insufficiency or illegibility of initial data; (3) the methods of the theory of artificial neural networks allowing to eliminate unauthenticity, insufficiency, a undefinitionality, inconsistency, incompleteness, inadequacy, inaccuracy of initial data for assessment and a categorization. Requirements to processes of assessment and a categorization in the conditions of uncertainty are formulated: efficiency, complexity, adequacy, reliability, objectivity, continuity, completeness, accuracy and openness. Requirements which are imposed to components of elimination of uncertainty of assessment and a categorization of semantic filling of IOs on the basis of use of methods of processing of incomplete, contradictory and indistinct knowledge are formulated: basic (on constant readiness,

stability, scalability (mobility portability), security of work, productivity) and additional (on modularity, functional selectivity, a generizability, virtualization, independence of programs of external devices, compatibility, openness, multitasking, a multi-user mode, security, profitability of work and also support of different file systems and standard formats of data). Execution of these requirements will allow intelligent scanners (qualifiers) quickly, fully (authentically), precisely and adequately to reveal signs and to counteract inappropriate, dubious and harmful information in digital network content.

6. The general approach and requirements imposed to components of adaptation and retraining of an analysis system of information objects including in the operation mode are developed. The purpose of the approach consists in development and the analysis of architecture of a system of classification of inappropriate, dubious and harmful information on a set of classification signs with use of text information of the websites, HTML data given about structure of documents and URL data using methods of automated classification of websites and their separate pages with use of methods of machine learning. The feature of the offered approach consists in hierarchical creation of the decisive rule in the form of a set of basic qualifiers and the top level qualifier integrating their initial data. The offered approach allows us to minimize the time expenditure connected with change of types of the analyzed initial data, modifying not the system of classification entirely, and only a separate subset of basic qualifiers, the corresponding aspect qualifier and the final qualifier, thereby adapting the available solution for the changed requirements. An approach proves three-level hierarchical architecture of classification of contents of web pages. At the first level binary classifiers which carry out check of belonging of the analyzed object to a certain category are placed. The second level includes the qualifiers intended for determination of aspects of the web page. At the third level the final qualifier is located. The initial data sets presented in the form of files of HTML pages are processed by a component of extraction of signs in the beginning. The feature sets created by it components are located in DDB of IO. Requirements to components on indicators of efficiency, scalability and reliability are proved.

7. The general approach and requirements imposed to components of development and choice of measures of counteraction of inappropriate, dubious and harmful information are developed. Approach allocates the ways and funds applicable for protection against information, and mechanisms of protection against information operating in the Russian Federation. The main objects and subjects interacting within this approach are defined: IO, information source, channel of distribution of information, target audience, subject of influence of information. Measures of counteraction are entered for protection of the subject of influence of information against harmful influence of information. They depend on classes of all listed subjects and objects and a class of information provided to IO (inappropriate, dubious and harmful). For target audience as the subject of influence of information three-level classification is offered. Channels of distribution of information are defined and their two-level classification is offered. Measures for protection against inappropriate, dubious and harmful information are classified by a way of protection in the beginning, and then by ways of their implementation. The main stages of counteraction on inappropriate, dubious and harmful information are: class definition of information, class definition of measures of protection, choice of optimum means / measures of protection. Functional requirements to components of development and the choice of measures of counteraction are: controllability of process; accounting of laws of the Russian Federation;

access restriction to inappropriate information; accounting of possible damage to citizens and state; process automation; respect for the rights of citizens to an information access.

8. The general approach and requirements imposed to components of implementation of visual interfaces for identification and counteraction of inappropriate, dubious and harmful information are developed. An approach, the called "visualization pipeline", consists in serial processing of data according to the following stages: (1) data analysis; (2) data filtering; (3) marking of data; (4) drawings of data. Besides, an approach correlates visualization models and a data structure. It allows us to implement the principle of recurrence when a set of models should be interchanged. Requirements to a visualization component from the process of visualization and architecture of this component are defined. From visualization the following demands are made: (1) possibility of identification of lie factor (wrong interpretation of a graphic metrics); (2) lack of chart junk (graphic elements which do not bear any information); (3) animation use. From architecture the following demands are made: (1) implementation of the concept of "visualization pipeline"; (2) standardization of data according to structure; (3) accounting ways of human-computer interaction with result of visualization; (4) accounting dependence of data volume from a way of their storage.

9. **The general architecture of perspective ISAPDNC for the purpose of detection and counteraction of inappropriate, dubious and harmful information** is developed. In the offered architecture components interact through DDB of IO. Data from components of preprocessing and elimination of incompleteness and inconsistency come to DDB. From DDB data come to components of elimination of incompleteness and inconsistency, adaptation and retraining, assessment and classification and counteraction. The component of ensuring efficiency and a component of visualization are backbone. The first of them provides implementation of methods and algorithms of parallel data processing in other components. The second component provides the visual interface of all components and display of results of their work in traditional and special visualization models. The test bed for pilot studies based on use of a VMware hypervisor and Hadoop and Spark platforms of distributed computing is developed for assessment of the architecture. The test dataset containing over 70 thousand records is created, each of which is a set of attributes of the web page of the Internet. Each web page belongs to one of 18 categories of inappropriate, dubious and harmful information. The received test dataset is approved at a solution of a problem of IO classification in two series of experiments: by one machine (the one-line mode) and on a cluster of the parallel working machines (the multithreaded mode). Results of experimental assessment showed that in both series the best results on training showed qualifiers on the basis of a support vector machine. At the same time it is established that the Spark system did not give a significant gain on training time that, apparently, is connected with insufficiently large volume of a test dataset.

By results of the project **8** publications in the editions indexed in Scopus base, **19** – in RISC base are made, and **7** certificates on the state registration of the computer programs is taken out.