Утверждено академическим руководителем

Образовательной Программы

«Системы больших данных» проф. Мальцева С.В.,

уровень образования: магистратура

одобрено Академическим Советом программы

14.10.2016

**1. New services based on open data/Новые сервисы на основе открытых данных**

Abstract: There should be presented new service based on open data concept, it should be described and analyzed with possible implementation (i.e. as mobile application). Literature review should show current existend solutions not just in Russia but all around the world. New services should not only utilize open data, but use open data as one part of the data source for the service.

Supervisor: Mikhail M. Komarov

**2.  D2D-communications business models/Бизнес-модели D2D**

Abstract: Device to Device communications should be analyzed and new approaches in terms of the business models should be proposed within this topic. There should be description of the communication technologies which are going to be used, storage of the data approach and description of the added value of the business model together with A.Osterwalder business model canvas description.

Supervisor: Mikhail M. Komarov

**3. New generation networks QoS for users /Качество представления услуг для пользователей в сетях нового поколения**

Abstract: Technologies for data transmission should be analyzed together with different approaches to measure quality of service for the end-users of the services which are provided through those networks. Positive and negative effects should be considered at the end for business-processe with the use of those technologies and dependancy of the proposed QoS metrics and business.

Supervisor: Mikhail M. Komarov

**4. Decentralized systems and analytical services/ Децентрализованные системы и аналитические сервисы**

Abstract: Decentralized approach should be presented and particular cases should be analyzed in terms of Big Data different types of data being key part of the business model. For instance, there might be analysis of the audio-services (like audio-chats, sharing music) as the crowd-services where obviously will be Big Data approach; or video (or pictures) sharing service - like live-streaming as a key point of the service and business.

Supervisor: Mikhail M. Komarov

**5. New nano-services and big data/ Новые сервисы нано-уровня и большие данные**

Abstract: Recent advancements in nano-technologies should be presented together with the analysis of the different levels of services in traditional business. Within this topic approaches to store the data, processing algorithms etc. can be analyzed and being core points of the thesis.

Supervisor: Mikhail M. Komarov

**6. Personal Data Protection during Big Data time depending on the context**

Abstract: Main goal of the project is to analyze and review existent approaches of personal data protection and define strategy to protect personal data depending on different types of context.

Main tasks:

1)    Review of existent approaches.

During this stage partners will overview existent approaches of personal data protection while working with Big Data.

2)    Review of existent and future contexts.

3)    During this stage partners will overview existent contexts and possible future context which will influence on personal data protection techniques.

4)    Development of the approach to protect data depending on different types of context.

During this stage partners will develop joint approach to protect personal data with dependency on the context using different anonymization techniques.

5)    Development of analytical questionnaire for further analysis.

During this stage partners will develop a questionnaire for survey which further will be implemented and analysis techniques which should be used to analyze the feedback from the companies on the proposed approach of personal data protection.

6)    Implementing survey among French companies working with personal data.

During this stage questionnaire will be sent to the companies and their responses will be analyzed.

7)    Conclusions and open issues.

During this stage partners will conclude the results of the project and describe open issues for further research.

As a main outcome from the project, there will be proposed approach of personal data protection depending on the context while working with Big Data with the responses from the industry.

Further, results of the project will be presented at the Internet Governance Forum (UN) as the personal data protection topic is extremely important while companies implementing data mining techniques and working with Big Data, where most of the data are acquired from the people.

Supervisor: Mikhail M. Komarov

7. **Strategic use and applications of big data technologies in higher education: the potential for big data to enhance the higher education sector in Russia**

Abstract: The availability of big data sets has stimulated the applications of big data technologies in higher education. Different issues in higher education could be solved in a more effective way due to big data technologies application. In the framework of this topic, the relevant examples of big data technologies application in higher education could be analyzed. As well as the challenges and difficulties that national higher education systems and particular universities face while applying the big data technologies. It could also cover the advantages and specificity of big data technologies application for different groups of users (students, academic and administrative staff).

Russian higher education is actively developing and leading universities are applying new technologies that permit to increase the quality of education and the effectiveness of university’s activities. Big data application is not a common practice for Russian universities. The results of foreign research approve that the application of big data technologies could enhance the higher education. In the framework of the topic it’s proposed to study the practices of different universities and national education systems of big data technologies application and formulate the proposal on how these practices could be used to enhance the higher education sector in Russia.

References:

1. Daniel B. (2015) Big Data and analytics in higher education: Opportunities and challenges // British Journal of Educational Technology, 46 (5), 904-920.
2. Ellaway R.H., Pusic M.V., Galbraith R.M., Cameron T. (2014) Developing the role of big data and analytics in health professional education // Medical Teacher, 36 (3), 216-222.
3. Johnson J. A. (2014) The Ethics of Big Data in Higher Education // International Review of Information Ethics, 07.
4. Krishnaveni S., Satheesh A., Kannan E. (2015) Review of big data on student information for finding the uncertainty in higher education enrollment // International Journal of Grid and High Performance Computing, 7 (4), 21-32.
5. Picciano, A.G. (2012) The evolution of big data and learning analytics in American higher education // Journal of Asynchronous Learning Network, 16 (3), 9-20.
6. Sang Q. (2013) Application of Big Data in Higher Education // Journal of Nantong Textile Vocational Technology College, 02.
7. Tulasi B. (2013) Significance of Big Data and Analytics in Higher Education //International Journal of Computer Applications, 68(14), 23–25.

# Supervisor: Alisa V. Melikyan

**8. Virality of modern business models**

Abstract: Virality is one of the modern properties of successful business models. It describes the positive effects while social media marketing and enables users massively try and share positive references about the product or service. Recent success of popular apps such as MSQD, PRISM, Pokemon GO and many others demonstrate how important is the integration between information systems and psychology. Despite of already existing competitors, some researchers claim that the virality was the decisive factor for their success, its spread looked more like a flash-mob than a planned marketing campaign.

In this course work student is expected to integrate case study material from various (mostly business) literature and create a set of recommendations for a start-up.

This hottest topic lays a good background for further master thesis and PhD dissertation worldwide.  It gives lots of tacit knowledge that one could apply starting own company or promoting products or services.

References:

1. <https://books.google.ru/books?hl=ru&lr=&id=z_VBKjscPLkC&oi=fnd&pg=PP2&dq=Virality+of+modern+business+models&ots=IxqpCqXbk0&sig=0ri01YyXfU1h24flGkli22xi7Eo&redir_esc=y#v=onepage&q&f=false>
2. <https://www.amazon.com/Invisible-Influence-Hidden-Forces-Behavior/dp/1476759693/ref=as_li_ss_tl?ie=UTF8&linkCode=sl1&tag=jbsite-20&linkId=11406899823cffd33264b6d0ea568c6f>
3. <https://www.amazon.com/Contagious-Things-Catch-Jonah-Berger/dp/1451686579/>
4. <http://jonahberger.com/>

# Supervisor: Nikolay S. Kazantsev

**9. Creation of business process models of the leading Russian aircraft carrier**

Abstract: Business process models describe logic of corporate performance and integrate several areas inc. enterprise mission, goals and objectives, 4 layers business and IT processes and IT-infrastructure. Airline industry is one of the most suitable for process management due to big amount of repetitive standardized operations and strict regulatory control. In this work a student will get practice in business process modelling (ARIS, Vision), creating presentations for customer (leading Russian airline carrier) and process analytics.

This topic is good for hands-on application for knowledge in process management, enterprise architecture and creates foundation for master thesis.

References:

1. Stavenko, Yulia, Nikolay Kazantsev, and Alexander Gromoff. "Business process model reasoning: from workflow to case management." Procedia Technology 9 (2013): 806-811.
2. <http://www.sciencedirect.com/science/article/pii/S096969970300084X>
3. <http://search.proquest.com/openview/01549212fb541abece8bed7a9f117b9e/1?pq-origsite=gscholar>
4. <http://search.proquest.com/openview/01549212fb541abece8bed7a9f117b9e/1?pq-origsite=gscholar>

# Supervisor: Nikolay S. Kazantsev

**10. Intercultural aspects of Big Data Processing Information Systems adaptation**

Abstract: During integration of any Information System (IS) there are often disagreements between clients of different levels (employees, managers) and external consultants who facilitate the job.  Very often they occur due to different levels of information culture, technology understanding or even intercultural aspects. While integration of newly appeared Big data systems, consultants need to take into consideration all these factors to decrease negative risks of the project failture. In this work, student is expected to work on an international World IT Project, that is executed in Russia till December to investigate the peculiarities of IS integration in Moscow. Her/His main focus would be on large-scale enterprise running Big Data processing systems.

This topic gives a good chance to get experience in the international scientific work, to perform a survey and a group of interviews, grab contacts both from international academia and Russian industry. It lays also foundation for a Master thesis.

References:

1. <http://dl.acm.org/citation.cfm?id=2367572>
2. <http://ieeexplore.ieee.org/document/6612229/>
3. <https://books.google.ru/books?hl=ru&lr=&id=AeYBavcZ9_UC&oi=fnd&pg=PP1&dq=Intercultural+aspects+Big+Data+Information+Systems&ots=Mt6pEo_765&sig=d_fqlHyes6UrM6lPbIP7emdGF0s&redir_esc=y#v=onepage&q&f=false>
4. <http://www.worlditproject.com/>

# Supervisor: Nikolay S. Kazantsev

**11. Big data analysis in industry N**

Abstract: Despite of all known characteristics of Big Data Systems (BDS) , they tend to be applied differently in various industries. Banking and Telecommunications, Public administration and Higher Education Institutions - all of those diverse industries have peculiar requirements for Big data Systems processing and application. The task for student here would be to compare the application of BDS in existing projects (mainly using academia and business literature) and to create an integrated table with requirements from business and delivered results. Student may also use data from a company he/she works for, this however might require an additional permission from the corporate compliance group or signing an NDA.

This topic lays also foundation for a Master thesis.

References:

1. Anna, Kuraeva, and Kazantsev Nikolay. "Survey on Big Data Analytics in Public Sector of Russian Federation." *Procedia Computer Science* 55 (2015): 905-911.
2. <http://ieeexplore.ieee.org/document/7067026/?arnumber=7067026>

# <http://www.sciencedirect.com/science/article/pii/S2213846313000114>

# <http://dl.acm.org/citation.cfm?id=2463712>

# Supervisor: Nikolay S. Kazantsev

**12. Technological trends (blockchain, uberisation, cloud sourcing and so on) and its influence on business architectures**

Abstract: Various technological trends shape business models and promotes flexibility and accessibility of services, product customization and many other advantages for an end-user. This topic was obligatory for every student in the form of Essay 1 in Enterprise Architecture Adaptation course. A student is expected to extend this essay, to fill it with more practical details and to work deeply on a chosen trend.

This topic lays also foundation for a Master thesis.

References:

1. Komarov, Mikhail, Nikita Konovalov, and Nikolay Kazantsev. "How Internet of Things Influences Human Behavior Building Social Web of Services via Agent-Based Approach." *Foundations of Computing and Decision Sciences*41.3 (2016): 197-210.
2. <http://www.sciencedirect.com/science/article/pii/S0024630109000594>

# Supervisor: Nikolay S. Kazantsev

**13. Analysis of complex system development based on fuzzy – cognitive mapping**

Abstract: For decision making on governance of complex socio- economic system, the key roles play a problem situation structuring and the finding-out significant factors of external environments of system. This process mostly depends on expert experience and available information analysis. If we can represent the system by formal model, the process of observing, analyzing of system development and problem identification can be considered as a systematic activity for monitoring of significant changes in external and internal environment, for observing of expert forecasts and analysis of their actuality based on monitoring of factor changes. Based on such systematically activity the strategy correction in holistic control cycle (goal-setting based on external environment prediction, analysis of existing system sate and finding-out a trajectory of strategy goals achievements, planning direction of tactic activities, implementation of plan and feed-back analysis for further strategic step) can be done depend on monitoring of significance changes in terms of influence on strategic goals achievements. With modern tools It seems to construct a supporting system that almost automatically clarify knowledge about problem, recommend significant variables. With processes automation and information development of society we got to big data set, produced by enterprise information system, internet. However the problem of lack information has not been solved. We got another problem with skewness in data, with unintegrated data and with quality of data.

Despite on rapidly growth of data volume describing system and processes and technological and analytical tools for its analysis, the problems of informayion environment have not lost their relevance and just have been transformed. Situations of application of the given approach are related to ill-structured  due to one of listed reasons:

* Lack of information and understanding about significant factors.
* Inhomogeneity of information about a situation connected with differences in quality and quantity of information about situation hampering the application of typical statistical methods of analysis for reveal factors determining situation. (for instance, political, economic, technological and so on).
* The large number of information sources and large stream of information from each source. The typical bid data tools applied for a situation monitoring allow to reveal thousands factors and interconnections between them out of millions but only part of them could use to analysis and simulation of a situation development.

In case of a new situation experts are main source of knowledge about influencing factors and structure of them interrelations. Big data tools could help to connect a situation model developing from casual beliefs of situation experts with variety of data sources characterizing the changing of factors specified by experts System for support of monitoring and analysis of such situations includes:

* module of formation of an expert network, chatbots system for the extraction of factors and relationships and collecting relevant factors;
* module for search information sources with estimation their relevance to a situation;
* module of intelligent agents for search of information about factors( subjects, events, newsbreaks), for interrelations extraction and the followed evaluation of level of factors change;
* module of knowledge visualization about a situation in the form of dashboards, including the visualization of the structure, dynamics of the significant factors and the forecast.

Cognitive modeling module allows optimizing the control action on a number of factors on the basis of the inverse problem solution on the cognitive box. The inverse problem solution allows to quickly picking a set of control actions on the problem situation, which will ensure optimal achievement of the set goals.

References:

1. Avdeeva Z., Kovriga S. (2008). Cognitive Approach in Simulation and Control. Proceedings of the 17th IFAC World Congress, Seoul, Korea, July 6-11, pp. 1613- 1620.
2. Avdeeva Z., Kovriga S. [Diagnosing of The Problematic Situation in Manufacturing System Development Based-On-Cognitive Map](https://publications.hse.ru/view/138532820) // Manufacturing Modelling, Management, and Control. 2013. Vol. 7, P.1. P. 964-968.
3. Gubanov, D., Korgin, N., Novikov, D., Raikov, A. (2014). E-Expertise: Modern Collective Intelligence, Springer. Series: Studies in Computational Intelligence, Vol. 558, XVIII, 112 p.
4. Kahraman, Cengiz et al. (2008) Fuzzy Multi-Criteria Decision Making -Springer. Pp.380
5. [Komarov M. M.](https://www.hse.ru/en/org/persons/14587151), Avdeeva Z. [Customer experience management for smart commerce based on cognitive maps](https://publications.hse.ru/view/161776450) // Annals of Data Science. 2016

# Supervisor: Zinaida K. Avdeeva

**14. Smart education environment**

Abstract: In order to develop custom approach in education in an modern education environment (EE), including off-line information sources of an university campus and valuable addition sources from internet, it is necessary to set several modules (stages) for composition of most optimal model of education for each student. In the first stage of interaction with digital educational platform, it would be most appropriate to conduct criterion-oriented test, which would not only determine cultural-cognitive profile, but also collect information about other basic characteristics, such as: motivation, educational background, informational and communication technology skills, professional interests. However, it is necessary to take into consideration the possible dynamic nature of before mentioned variables. It would be rational to retest an individual periodically, and make corrections to the selected education course and form of educational environment by monitoring and including information, knowledge, case studies from different on-line sources [2].

The smart learning system for support dynamic personal educational pathway in an modern EE should include adaptation functions [1]:

1. The functions for supporting student learning process in active electronic educational environment

* The mapping of personal study pathway including to support of the different level of selection (several functions);
* Planning of the educational process according to the choice;
* Execution and control of the strategy.

1. The functions of interactive support of tutor’s work

* Tools for detection and construction of systematic linkages between courses and competencies
* Tools for monitoring space of the course that is manifested as a relation between actual knowledge and the results of didactic processing of such knowledge for a certain specialty, which are synchronized with the applied aspect (practical utility).

3) The functions of interactive support based on several i-agents. In such case, electronic environment is built on some repository of all teaching materials or virtual links to them in the system of curriculum’s courses. Therefore, the necessary functionality of such information-analytical system based-on-i-agents for the dynamic formation of EEE will contain the following: search, surfing, recommendations, and navigation.

4) As for other users of the system, apart from the direct users of EEE who support the learning process, it is worth noting the possibility of participation of methodists and professional associations (or, in other words, "consumers" of graduates), which can enter new market position in the system, using "language" of competencies or to find the closest set of courses for optimal specialization or even correction of existing courses.

References:

1. Avdeeva Z., Taratuhina Y. V., Omarova N. O. Smart Educational Environment as a Platform for Individualized Learning Adjusted to Student's Cultural-Cognitive Profile, in: Smart Education and Smart e-learning. Switzerland : Springer International Publishing, 2015. P. 219-231.
2. Conati, C. (2009). Intelligent Tutoring Systems: New Challenges and Directions. Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, pp. 2-7
3. Smart Education and Smart e-learning. Switzerland : Springer International Publishing, 2015. P. 570.
4. B. Kopp, M. Matteuccib

E-tutorial support for collaborative online learning: An explorative study on experienced and inexperienced e-tutors//Computer &Education, 58 (2012), pp. 12–20

# Supervisor: Zinaida K. Avdeeva

**15. The development of self-organizing neural network model of the financial systems in the context of big data**

Abstract: A self-organizing map (SOM) or self-organising feature map (SOFM) is a type of [artificial neural network](https://en.wikipedia.org/wiki/Artificial_neural_network) (ANN) that is trained using [unsupervised learning](https://en.wikipedia.org/wiki/Unsupervised_learning) to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a *map*. Self-organizing maps are different from other artificial neural networks as they apply [competitive learning](https://en.wikipedia.org/wiki/Competitive_learning) as opposed to error-correction learning (such as [backpropagation](https://en.wikipedia.org/wiki/Backpropagation) with [gradient descent](https://en.wikipedia.org/wiki/Gradient_descent)), and in the sense that they use a neighborhood function to preserve the [topological](https://en.wikipedia.org/wiki/Topology) properties of the input space.

This makes SOMs useful for [visualizing](https://en.wikipedia.org/wiki/Scientific_visualization) low-dimensional views of high-dimensional data and this case completely suites for analysis of multidimensional financial data. On top of everything else self-organizing neural network refers to the terms “synergy” and “synergetics”. This gives us ability to work with nonlinear models, which are more effective in financial prediction rather than linear ones.

References:

1. T. Kohonen, Self-Organization and Associative Memory. Springer, Berlin, 1984.
2. Von der Malsburg, C (1973). "Self-organization of orientation sensitive cells in the striate cortex". Kybernetik. 14: 85–100.
3. Kohonen, Teuvo; Honkela, Timo (2007). ["Kohonen Network"](http://www.scholarpedia.org/article/Kohonen_network). Scholarpedia.

# Supervisor: Nikolay V. Markov

**16. Modeling and implementation of Elliot waves to Big Data time series**

Abstract: The Elliott Wave Principle is a form of [technical analysis](https://en.wikipedia.org/wiki/Technical_analysis) that traders use to analyze financial market cycles and forecast [market trends](https://en.wikipedia.org/wiki/Market_trend) by identifying extremes in investor psychology, highs and lows in prices, and other collective factors. [Ralph Nelson Elliott](https://en.wikipedia.org/wiki/Ralph_Nelson_Elliott) , a professional accountant, discovered the underlying social principles and developed the analytical tools in the 1930s. He proposed that market prices unfold in specific patterns, which practitioners today call Elliott waves, or simply waves. Elliott stated that "because man is subject to rhythmical procedure, calculations having to do with his activities can be projected far into the future with a justification and certainty heretofore unattainable." The empirical validity of the Elliott Wave Principle remains the subject of debate. Try your scientific imagination and use the most actual mathematical tools for Big Data time series forecasting and analysis.

# Supervisor: Nikolay V. Markov

**17. Fuzzy logic tools in Big Data**

Abstract: Fuzzy logic is a form of [many-valued logic](https://en.wikipedia.org/wiki/Many-valued_logic) in which the [truth values](https://en.wikipedia.org/wiki/Truth_value) of variables may be any real number between 0 and 1, considered to be "fuzzy". By contrast, in [Boolean logic](https://en.wikipedia.org/wiki/Boolean_algebra), the truth values of variables may only be the "crisp" values 0 or 1. Fuzzy logic has been employed to handle the concept of partial truth, where the truth value may range between completely true and completely false. Furthermore, when [linguistic](https://en.wikipedia.org/wiki/Linguistic) variables are used, these degrees may be managed by specific (membership) functions.

These specific features of fuzzy logic give us a high range of possibilities in the field of Big Data. Develop your own fuzzy tools for Big Data compression, storage or other processes.

References:

1. [Zadeh, L.A.](https://en.wikipedia.org/wiki/Lotfi_A._Zadeh) (1968). "Fuzzy algorithms". Information and Control. 12 (2): 94–102.
2. Zadeh, L.A. (1965). "Fuzzy sets". Information and Control. 8 (3): 338–353.
3. Fernandez, A., Carmona, C.J., Jesus, M.J., Herrera, F. A View on Fuzzy Systems for Big Data: Progress and Opportunities. International Journal of Computational Intelligence Systems, Vol. 9, Supplement 1 (2016), 69-80

# Supervisor: Nikolay V. Markov

**18. Two-layer interval weighted graphs in assessing the market risks on big data**

Abstract: This scientific work is dedicated to the development, improvement and application of double layer interval weighted graphs (DLIG) for non-stationary time series forecasting. This model appears to be the universal and easy-to-use tool for modeling the non-stationary time series and forecasting. We observe the double layer version of the model because it’s the most representative way in the sense of main idea though you can add several layers more for different purposes. The first layer of the graph is based on empirical fluctuations of system and displays the most potential fluctuations of the system at the time of system training. The second layer of the graph as a superstructure of the first layer displays the degree of modeling error and it’s connected with the first layer nodes by edges. The second layer is the way of supervised training implementation with the aim of error minimization.

References:

1. Jingfei YM, Power system short-term load forecasting. Thesis for Ph. Degree. Germany, Darmstadt, Elektrotechnik und Informationstechnik der Technischen Universitat, p. 139.
2. Chen SM and Hsu CC, 2004. A new method to forecast enrollments using fuzzy time series. Int. Journal Applied Science and Engineering, Vol. 2, p. 234–244
3. Huarng K, 2004. Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets and Syst., Vol. 123, p. 387–394

# Supervisor: Nikolay V. Markov

**19. Dynamic analysis of big stream data using a financial platform**

Abstract: Data Stream Mining is the process of extracting knowledge structures from continuous, rapid data records. A data stream is an ordered sequence of instances that in many applications of data stream mining can be read only once or a small number of times using limited computing and storage capabilities. In many data stream mining applications, the goal is to predict the class or value of new instances in the data stream given some knowledge about the class membership or values of previous instances in the data stream. Machine learning techniques can be used to learn this prediction task from labeled examples in an automated fashion. Often, concepts from the field of incremental learning, a generalization of [Incremental heuristic search](https://en.wikipedia.org/wiki/Incremental_heuristic_search) are applied to cope with structural changes, on-line learning and real-time demands. In many applications, especially operating within non-stationary environments, the distribution underlying the instances or the rules underlying their labeling may change over time, i.e. the goal of the prediction, the class to be predicted or the target value to be predicted, may change over time. This problem is referred to as [concept drift](https://en.wikipedia.org/wiki/Concept_drift).

References:

1. Rutkowski, Jaworski, Pietruczuk and Duda: ["A New Method for Data Stream Mining Based on the Misclassification Error"](http://www.iisi.pcz.pl/~rutkowski/papers/TNNLS_2015.pdf), IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 5, pp. 1048-1059, 2015.
2. Shaker, Ammar and Lughofer, Edwin. ["Self-Adaptive and Local Strategies for a Smooth Treament of Drifts in Data Streams."](http://link.springer.com/article/10.1007%2Fs12530-014-9108-y), Evolving Systems, 5:(4), p. 239-257, 2014.
3. Klinkenberg, Ralf: Using Labeled and Unlabeled Data to Learn Drifting Concepts. In Kubat, Miroslav and Morik, Katharina (editors), Workshop notes of the IJCAI-01 Workshop on \em Learning from Temporal and Spatial Data, pages 16–24, IJCAI, Menlo Park, CA, USA, AAAI Press, 2001.

# Supervisor: Nikolay V. Markov

**20. Using Elasticsearch for data processing needs.**

Abstract: The work should describe Elasticsearch solution as a tool used to make queries throughout various data sets. Student must research the software, list its main features, provide analysis of advanced features and make an example of Elasticsearch usage for an analytical task chosen in collaboration with projects supervisor.

References:

1. <https://www.elastic.co/>
2. <https://www.kaggle.com/>

# Supervisor: Petr A. Baranov

**21. Log management and analysis with Logstash software as an example.**

Abstract: Student must research modern log management methods and issues. Provide description of common log gathering and storage methods. Logstash software should be used to illustrate management of vast amounts of logs using single software tool.

References:

1. <https://www.elastic.co/products/logstash>
2. <https://www.kaggle.com/>

# Supervisor: Petr A. Baranov

**22. Visualization of data analysis using Kibana**

Abstract: Student chooses several different datasets and defines analytical tasks for the data (in collaboration with project supervisor). Student must make an overview of data visualization tools that could be used to create an easy-to-understand representation of data analysis. The results must be visualized using Kibana plugin for Elasticsearch.

References:

1. <https://www.elastic.co/products/kibana>
2. <https://www.kaggle.com/>

# Supervisor: Petr A. Baranov

**23. Financial data visualization**

Abstract: The complexity of financial market data and the tasks involved in analyzing this data call for a visual tool that can amplify human cognition.

To help people gain insight into stock market data, a visual analytical tools are developed that analyze the historical price movements of publicly traded companies, that cluster similar data together, and that visualize the data onto a two-dimensional space using a machine learning algorithms.

Researchers and developers of financial data visualization tools combine learning algorithms and information visualization to exploit human perceptual ability to recognize patterns and derive a lot of information from visualization with little effort.

The objective is to study approaches to financial data visualization and compare tools available for financial data visualization and analysis.

https://d.adroll.com/cm/index/out https://d.adroll.com/cm/n/out

References:

1. Joel Joseph and Indratmo. Visualizing Stock Market Data with Self-Organizing Map. In: Proc. 26Th Int'l Florida Artificial Intelligence Research Society Conference. Copyright © 2013, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved. pp.488-491. ([pdf](https://www.google.ru/url?sa=t&rct=j&q=&esrc=s&source=web&cd=5&ved=0ahUKEwjklYijktDPAhWCFCwKHUsGClwQFghEMAQ&url=https%3A%2F%2Fwww.aaai.org%2Focs%2Findex.php%2FFLAIRS%2FFLAIRS13%2Fpaper%2Fdownload%2F5937%2F6123&usg=AFQjCNE6MOTn415Bd4WQKhVsy7sNlo4aqA&cad=rjt))
2. Krešimir Šimunić Visualization of Stock Market Charts. Semantic scholar.org. - <https://pdfs.semanticscholar.org/157a/1cb155ea34d33eacce0c870e8d6fd60a6c13.pdf>
3. Eric Greenbaum. #179: Visualizing Stock Market Data & Bringing VR to Wall Street with QuantVR. Voices of VR Podcast, July 30, 2015. - <http://voicesofvr.com/179-visualizing-stock-market-data-bringing-vr-to-wall-street-with-quantvr/>
4. Rui Ma. Designing interactive visualization methods for comparing multivariate stock market data over time. Diploma work. 2009. - [www.cvast.tuwien.ac.at/sites/default/files/thesis\_design.pdf](http://www.cvast.tuwien.ac.at/sites/default/files/thesis_design.pdf)
5. Jimmy Steinmetz. Pulling Stock Data for Visualization in Tableau. 06.19.2015 - <https://www.interworks.com/blog/jsteinmetz/2015/06/19/pulling-stock-data-visualization-tableau>
6. Data visualizations by U.S. Securities and Exchange Commission - <https://www.sec.gov/marketstructure/datavis.html>

# Supervisor: Petr B. Panfilov

**24. Big data analytics and modern industrial automation and control**

Abstract: Supervisory control and data acquisition (SCADA) is a system for [remote monitoring and control](https://en.wikipedia.org/wiki/Remote_monitoring_and_control) that operates with coded signals over communication channels (using typically one communication channel per remote station) (Wikipedia).

SCADA systems historically distinguish themselves from other [industrial control systems](https://en.wikipedia.org/wiki/Industrial_control_system) (ICS) by being large-scale processes that can include multiple sites, and large distances. These processes include industrial, infrastructure, and facility-based processes, as described below:

* [Industrial processes](https://en.wikipedia.org/wiki/Industrial_process) include those of manufacturing, production, [power generation](https://en.wikipedia.org/wiki/Power_generation), [fabrication](https://en.wikipedia.org/wiki/Fabrication_(metal)), and refining, and may run in continuous, batch, repetitive, or discrete modes.
* [Infrastructure](https://en.wikipedia.org/wiki/Infrastructure) processes may be public or private, and include [water treatment](https://en.wikipedia.org/wiki/Water_treatment) and distribution, wastewater collection and [treatment](https://en.wikipedia.org/wiki/Waste_water_treatment), [oil and gas pipelines](https://en.wikipedia.org/wiki/Fossil_gas_pipeline), [electrical power transmission](https://en.wikipedia.org/wiki/Electrical_power_transmission) and [distribution](https://en.wikipedia.org/wiki/Electric_power_distribution), [wind farms](https://en.wikipedia.org/wiki/Wind_farm), [civil defense siren](https://en.wikipedia.org/wiki/Civil_defense_siren) systems, and large communication systems.
* Facility processes occur both in public facilities and private ones, including buildings, airports, [ships](https://en.wikipedia.org/wiki/Ship#Today), and [space stations](https://en.wikipedia.org/wiki/Space_station). They monitor and control [heating, ventilation, and air conditioning](https://en.wikipedia.org/wiki/Heating,_ventilation,_and_air_conditioning) systems (HVAC), [access](https://en.wikipedia.org/wiki/Access_control), and [energy consumption](https://en.wikipedia.org/wiki/Efficient_energy_use).

With the commercial availability of [cloud computing](https://en.wikipedia.org/wiki/Cloud_computing), SCADA systems have increasingly adopted [Internet of Things](https://en.wikipedia.org/wiki/Internet_of_Things) technology to significantly reduce infrastructure costs and increase ease of maintenance and integration. As a result, SCADA systems can now report state in near real-time and use the horizontal scale available in cloud environments to implement more complex control algorithms than are practically feasible to implement on traditional [programmable logic controllers](https://en.wikipedia.org/wiki/Programmable_logic_controller).

This decentralization of data also requires a different approach to SCADA than traditional PLC based programs. When the data comes from a disparate mix of sensors, controllers and databases (which may be local or at varied connected locations), the typical 1 to 1 mapping becomes problematic. A solution to this is [data modeling](https://en.wikipedia.org/wiki/Data_modeling), a concept derived from object oriented programming.

In a data model, a virtual representation of each device is constructed in the SCADA software. These virtual representations (“models”) can contain not just the address mapping of the device represented, but also any other pertinent information (web based info, database entries, media files, etc.) that may be used by other facets of the SCADA/IoT implementation. As the increased complexity of the Internet of Things renders traditional SCADA increasingly “house-bound,” and as communication protocols evolve to favor platform-independent, service-oriented architecture (such as [OPC UA](https://en.wikipedia.org/wiki/OPC_UA)), it is likely that more SCADA software developers will implement some form of data modeling.

The objective is to study data modeling/analysis/visualization approaches to distributed SCADA/IoT implementations.

References:

1. J. Russel. ["A Brief History of SCADA/EMS (2015)"](http://scadahistory.com/)
2. Real time data visualization essentials. White papers by B-Scada - <https://scada.com/Content/Whitepapers/Real-Time%20Data%20Visualization%20Essentials.pdf>
3. [Kitamura, Misayo](http://adsabs.harvard.edu/cgi-bin/author_form?author=Kitamura,+M&fullauthor=Kitamura,%20Misayo&charset=ISO-8859-1&db_key=PHY); [Kojima, Taizo](http://adsabs.harvard.edu/cgi-bin/author_form?author=Kojima,+T&fullauthor=Kojima,%20Taizo&charset=ISO-8859-1&db_key=PHY); [Nishida, Shogo](http://adsabs.harvard.edu/cgi-bin/author_form?author=Nishida,+S&fullauthor=Nishida,%20Shogo&charset=ISO-8859-1&db_key=PHY). SCADA Data Visualization Using Equipment Graphs. IEEJ Transactions on Electronics, Information and Systems, Volume 126, Issue 6, pp. 788-796 (2006).
4. [How The "Internet Of Things" Is Turning Cities Into Living Organisms](http://www.fastcompany.com/biomimicry/how-the-internet-of-things-is-turning-cities-into-organisms/) Retrieved September 16, 2013.
5. Melissa Topp. Data historians and HMI: The foundation of big data analytics. Control Engineering online magazine, 05.07.2015. - <http://www.controleng.com/single-article/data-historians-and-hmi-the-foundation-of-big-data-analytics/7e50ad362333216b8f9a5e7e7c3db722.html>

# Supervisor: Petr B. Panfilov

**25. The analysis of the approaches to designing natural language interfaces for interaction with Linked Open Data /** Анализ подходов к разработке естественно-языковых интерфейсов для взаимодействия с Системой взаимосвязанных открытых данных

Abstract: During last decade, a big data system has emerged called Linked Open Data (LOD). This system can be imagined as a huge marked graph. It is composed of elementary graphs representing the triples of the language RDF (Resource Description Framework). For interaction with LOD, a special request language SPARQL has been developed. The role of SPARQL for LOD is similar to the role of the language SQL for the interaction with relational databases. For mastering the main constructions of SPARQL, it is necessary to have a background in informational languages. That is why it is important to develop natural language interfaces for interaction with LOD. The idea of the studies to be analysed is to develop a transformer from a concrete natural language (for instance, English, German, or Russian) into the language SPARQL.

References

1. Hladky, D. and Maltseva, S.V. Linked Data Paradigm for Enterprises: Information Integration and Value Chain // Business Informatics. A scientific-practical journal of National Research Univ. Higher School of Economics (HSE). 2013, No. 2 (24), p. 3-12.
2. Resource Description Framework (RDF): Concepts and Abstract Syntax, Graham Klyne and Jeremy J. Carroll, Editors. W3C Recommendation, 10 February 2004  (2004) \\ \url{http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/.  Latest version available at <http://www.w3.org/TR/rdf-concepts/>}
3. RDF Vocabulary Description Language 1.0: RDF Schema , Dan Brickley and R.V. Guha, Editors. W3C Recommendation, 10 February 2004 (2004) \\ \url{http://www.w3.org/TR/2004/REC-rdf-schema-20040210/.
4. **Wang C., Xiong M., Zhou Q., Yu Y.** PANTO: a portable NL-interface to ontologies *4th European Semantic Web Conference Proceedings. Springer,* 2007, pp. 473-487.
5. **Fomichov, V.A.: Semantics-Oriented Natural Language Processing: Mathematical Models and Algorithms. Springer, New York, Dordrecht, Heidelberg (2010a)**
6. **Fomichov, V.A.: Theory of K-representations as a Comprehensive Formal Framework for Developing a Multilingual Semantic Web. Informatica. An International Journal of Computing and Informatics  34 (3), 387-396 (2010b).**

# Supervisor: Vladimir A. Fomichov

**26. The analysis of the semantics-oriented computational approaches to discovering significant events' descriptions in natural language texts/** Разработка семантически-ориентированных компьютерных подходов к обнаружению описаний важных событий в естественно-языковых текстах

Abstract: The term paper should analyze the main approaches to the development of algorithms for discovering significant business events' descriptions in natural language texts. Special attention should be paid to the used formal means for representing the input, intermediate, and output data. The output data are semantic representations of the analyzed fragments of natural language texts.

References

1. **Lei Y., Uren V., Motta E**. SemSearch: A Search Engine for the Semantic Web. S. Staab and V. Svatek (Eds.), EKAW 2006 - Proc. 15th Intern. Conf. on Knowledge Enguineering and Knowledge management, 2-6 Oct. 2006, Podebrady, Czech Republic. Lecture Notes in Artificial Intelligence. Berlin, Heidelberg: Springer, 2006, vol. 4248, pp. 238-245.
2. **Fomichov V. A., Kirillov A. V.** A Formal Model for Constructing Semantic Expansions of the Search Requests about the Achievements and Failures. *Ramsay A., G. Agre G. (Eds). Artificial Intelligence: Methodology, Systems, and Applications. Lecture Notes in Computer Science.* Berlin, Heidelberg: Springer, 2012, vol. 7557, pp. 296–304.
3. **Aggarwal N., Polajnar T., Buitelaar P**. Cross-Lingual Natural Language Querying over the Web of Data Metais E., Meziane  F., Saraee M., Sugumaran V., Vadera S. (eds.). Natural Language Processing and Information Systems. 18th Intern. Conference on Applications of Natural Language to Information Systems, NLDB 2013, Salford, UK, June 2013, Proceedings. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2013, vol. 7934, pp. 152-163.
4. **Fomichov, V.A.: Semantics-Oriented Natural Language Processing: Mathematical Models and Algorithms. Springer, New York, Dordrecht, Heidelberg (2010a)**
5. **Fomichov, V.A.: Theory of K-representations as a Comprehensive Formal Framework for Developing a Multilingual Semantic Web. Informatica. An International Journal of Computing and Informatics  34 (3), 387-396 (2010b).**
6. **Fomichov, V.A.: A Broadly Applicable and Flexible Conceptual Metagrammar as a Basic Tool for Developing a Multilingual Semantic Web. In: Metais, E., Meziane, F., Saraee, M., Sugumaran, V., Vadera, S. (eds.) NLDB 2013. LNCS, Vol. 7934, 249-259. Springer, Heidelberg (2013)**
7. **Fomichov, V.A.: SK-Languages as a Comprehensive Formal Environment for Developing a Multilingual Semantic Web. In: Decker, H., Lhotska, L., Link, S., Spies, M, Wagner, R. R. (eds.) Database and Expert Systems Applications, 25th Intern. Conference, DEXA 2014, Munich, Germany, September 1-4, 2014, Proceedings, Part I, LNCS, Vol. 8644, 394-401. Springer, Cham, Heidelberg (2014)**

# Supervisor: Vladimir A. Fomichov

**27. The analysis of the approaches to designing transformers of natural language texts into the collections of OWL-expressions**/Анализ подходов к проектированию преобразователей естественно-языковых текстов в наборы выражений языка OWL

Abstract: During last decade, the researchers in various scientific-technical centres in the world have created a big family of Web-based ontologies (knowledge bases) in many thematic domains. For representing knowledge pieces, they mainly used the language of developing ontologies OWL (Ontology Web Language), created under the framework of very large-scale Semantic Web project. The volume of efforts for constructing ontologies is so high that we need to have a way of automatically doing this. This way is automatic extraction of knowledge from texts in natural language (the union of English, Russian, German, French, Chinese, Japanese, and many other languages). The objective of term paper is to analyse the current situation and the main trends in this field.

References

1. OWL Web Ontology Language Semantics and Abstract Syntax , Peter F. Patel-Schneider, Patrick Hayes, and Ian Horrocks, Editors. W3C Recommendation, 10 February 2004  (2004)\\ \url{http://www.w3.org/TR/2004/REC-owl-semantics-20040210/.  Latest version available at <http://www.w3.org/TR/owl-semantics/>}
2. OWL Web Ontology Language Overview , Deborah L. McGuinness and Frank van Harmelen, Editors. W3C Recommendation, 10 February 2004  (2004)\\ \url{http://www.w3.org/TR/2004/REC-owl-features-20040210/.  Latest version available at <http://www.w3.org/TR/owl-features/>}
3. Heflin J OWL Web Ontology Language Use Cases and Requirements [электронныйресурс] W3C Recommendation – 2004 -  URL: <http://www.w3.org/TR/2004/REC-webont-req-20040210>.
4. Chang-Shing Leea, Yuan-Fang Kaob, Yau-Hwang Kuob, Mei-Hui Wanga Automated ontology construction for unstructured text documents - Data & Knowledge Engineering Volume 60, Issue 3, March 2007, Pages 547–566.
5. Faria C., Serra I , Girardi R.  A domain-independent process for automatic ontology population from text - Science of Computer Programming Volume 95, Part 1, 1 December 2014, Pages 26–43.
6. Pantel P. and Pennacchiotti M.  Automatically Harvesting and Ontologizing Semantic Relations. In Paul Buitelaar and Philipp Cimiano (Eds.) Ontology Learning and Population: Bridging the Gap between Text and Knowledge - Selected Contributions to Ontology Learning and Population from Text. 2008ISBN: 978-1- 58603-818-2.
7. **Fomichov, V.A.: Semantics-Oriented Natural Language Processing: Mathematical Models and Algorithms. Springer, New York, Dordrecht, Heidelberg (2010a)**

# Supervisor: Vladimir A. Fomichov

**28. The analysis of the approaches to designing transformers of the collections of OWL-expressions into natural language texts/**Анализ подходов к проектированию преобразователей наборов выражений языка OWL в  естественно-языковые тексты

Abstract: Suppose that an IT-company or  a research group wants to conclude a contract for solving a technical task on the basis of interaction with an ontology (a knowledge base). That is why it is  highly important to quickly become aware of the subsystem of knowledge presented in the ontology. A transformer of the mentioned kind is able to help to quickly understand the situation with relevant knowledge presented in an ontology.

# Supervisor: Vladimir A. Fomichov

**30. Start-up and configuration BOINC-server and publics computing for big radioastronomical data.**

Abstract: In 2012 the multi-beam feed array became operational on the BSA FIAN radio telescope [1,2,3]. Today it is capable of 24-hour observation using 96 beams in declination in the sky of -8 to +42 degrees (aboutly 40% of the sky) in the 109-111.5 MHz frequency band. The number of frequency bands range from 6 to 32, while the time constants range from 0.1 to 0.0125 sec. While receiving in 32 band mode (plus one common band) with a time constant of 12.5 ms (80 times per second), 33x96x80 four-byte data samples are produced per second, which equates to a daily data production of 87.5 gigabytes (up to 32 terabytes yearly, to this time 70 TB in archives). This data is an enormous opportunity for both short and long-term monitoring of various classes of radio sources (including radio transients), space weather and the Earth's ionosphere monitoring, search for different classes of radio sources, etc.

Ultimately, the multi-beam observations of the BSA of LPI telescope will take advantage of the wide-field survey capabilities to enable the discovery and investigation of variable and transient phenomena from the intra- to the extra- galactical, including flare stars, intermittent pulsars, X-ray binaries, magnetars, extreme scattering events, interstellar scintillation, radio supernovae and orphan afterglows of gamma-ray bursts.

But at this moment we have not good instrument for stream computing of these big data. We need in launch the streaming data on various types of high-performance computing systems, including to create a public system of distributed computing for thousands of users on the basis of BOINC technology. So, thus instrument can be the BOINC technology [4]. You must start-up, configuring and run BOINC-server [5] and create client part for it on base some computing and astronomical algorithms (sorting, correlation, Fourier analytics etc. also as some astronomical algorithms). Need skills from You: C++ (preferably also in a Qt environment), Linux, SQL (MYSQL and Postgresql preferable), good skills also will Perl or/and Python.

The BOINC client for astronomical data from the monitoring survey of the big part of entire sky almost have not analogies.

**References**

1. Oreshko V.V. et al : 2012, Transactions of the Institute of Applied Astronomy (Russia), v. 24, p. 80
2. <http://astro.prao.ru/> - on-line database from daily monitoring of the radio sky in the 109-111.5 MHz (from 7 July 2012 to 20 October 2013).
3. Samodurov V. A., Rodin A.E., Kitaeva M. A., Isaev E., Dumsky D. V., Churakov D.D., Manzyuk M.O.: The daily 110 MHz sky survey (BSA FIAN): online database, science goals data processing by distributed computing. // Труды ХVII международной конференции «Аналитика и управление данными в областях с интенсивным использованием данных» (“Data Analytics and Management in Data Intensive Domains” (DAMDID)). Обнинск: НИЯУ МИФИ, 2015. P. 127-128.
4. <http://boinc.berkeley.edu/wiki/User_manual>
5. <https://boinc.berkeley.edu/trac/wiki/ServerIntro>

# Supervisor: Vladimir A. Samodurov

**31. The creation www-site for big data project on base radioastronomy data.**

Abstract: In 2012 the multi-beam feed array became operational on the BSA FIAN radio telescope [1,2,3]. Today it is capable of 24-hour observation using 96 beams in declination in the sky of -8 to +42 degrees (aboutly 40% of the sky) in the 109-111.5 MHz frequency band. The data receiving in 32 band mode (plus one common band) with a time constant of 12.5 ms (80 times per second), 33x96x80 four-byte data samples are produced per second, which equates to a daily data production of 87.5 gigabytes (up to 32 terabytes yearly, to this time 70 TB in archives). This data is an enormous opportunity for both short and long-term monitoring of various classes of radio sources (including radio transients), space weather and the Earth's ionosphere monitoring, search for different classes of radio sources, etc. For example, today more than 20 new pulsars from this data have been discovering (from 2,5 thousands in total), that one from best results for pulsar found commands in the word. At this moment tested different ways of data processing.

But at this moment we have not good www site for our big data science project. On site [2] we have only simple pattern for one from many needed functions of future site. On this site is needed:

- base information about project on English and Russian;

- some administrative tolls for adding information;

- some cgi scripts for visualisation data from database of proect;

- communication tolls for science moderators and volunteers of project (forum, for example);

- and so one (Your task is construct and propose good mechanism for public big data science project).

Need skills from You: html, Linux, SQL (Postgresql preferable), good skills also with some scripts language (for example Perl or/and Python).

You will work in close contact with other colleagues from this project and authors close course work subject.

**References**

1. Oreshko V.V. et al : 2012, Transactions of the Institute of Applied Astronomy (Russia), v. 24, p. 80
2. <http://astro.prao.ru/> - on-line database from daily monitoring of the radio sky in the 109-111.5 MHz (from 7 July 2012 to 20 October 2013).
3. Samodurov V. A., Rodin A.E., Kitaeva M. A., Isaev E., Dumsky D. V., Churakov D.D., Manzyuk M.O.: The daily 110 MHz sky survey (BSA FIAN): online database, science goals data processing by distributed computing. // Труды ХVII международной конференции «Аналитика и управление данными в областях с интенсивным использованием данных» (“Data Analytics and Management in Data Intensive Domains” (DAMDID)). Обнинск: НИЯУ МИФИ, 2015. P. 127-128.

# Supervisor: Vladimir A. Samodurov

**32. The creation administrative and automation tools for database of big data project on base radioastronomy data.**

Abstract: In 2012 the multi-beam feed array became operational on the BSA FIAN radio telescope [1,2,3]. Today it is capable of 24-hour observation using 96 beams in declination in the sky of -8 to +42 degrees (aboutly 40% of the sky) in the 109-111.5 MHz frequency band. The data receiving in 32 band mode (plus one common band) with a time constant of 12.5 ms (80 times per second), 33x96x80 four-byte data samples are produced per second, which equates to a daily data production of 87.5 gigabytes (up to 32 terabytes yearly, to this time 70 TB in archives). This data is an enormous opportunity for both short and long-term monitoring of various classes of radio sources (including radio transients), space weather and the Earth's ionosphere monitoring, search for different classes of radio sources, etc. For example, today more than 20 new pulsars from this data have been discovering (from 2,5 thousands in total), that one from best results for pulsar found commands in the word. At this moment tested different ways of data processing.

For these big data is constructed database on base of Postgresql [4]. But this database have not good administrative and automation tools for addition and pre-processing our big data science project. So we have in our data base only for 2012-2013, we not add last data and dada for big regime of observation. We need in systematic information from observational servers, also as preliminary preparation of data for stream processing (other course work). The science objesct table is need for constructing, it is need constructed some auto information for site of project (other course work) and so one.

Need skills from You: SQL (Postgresql preferable), Linux, html, good skills also with some scripts language (for example Perl or/and Python) and C++.

**References**

1. Oreshko V.V. et al : 2012, Transactions of the Institute of Applied Astronomy (Russia), v. 24, p. 80
2. <http://astro.prao.ru/> - on-line database from daily monitoring of the radio sky in the 109-111.5 MHz (from 7 July 2012 to 20 October 2013).
3. Samodurov V. A., Rodin A.E., Kitaeva M. A., Isaev E., Dumsky D. V., Churakov D.D., Manzyuk M.O.: The daily 110 MHz sky survey (BSA FIAN): online database, science goals data processing by distributed computing. // Труды ХVII международной конференции «Аналитика и управление данными в областях с интенсивным использованием данных» (“Data Analytics and Management in Data Intensive Domains” (DAMDID)). Обнинск: НИЯУ МИФИ, 2015. P. 127-128.
4. <https://www.postgresql.org/docs/9.5/static/docguide.html>

# Supervisor: Vladimir A. Samodurov

**33. Big Data in Bioinformatics and Healthcare**

# Supervisor: Vasily V. Kornilov

**34. Big Data in Personalized Medicine**

# Supervisor: Vasily V. Kornilov

**35. Digital Business Transformation**

# Supervisor: Vasily V. Kornilov

**36. Feasibility Study on the Use of Big Data Technologies (for a company or organization)**

# Supervisor: Vasily V. Kornilov

**37. Modern Information Society: Information overload and Information explosion problems**

# Supervisor: Vasily V. Kornilov

**38. Approaches to the Modeling of Large Social Networks**

# Supervisor: Olga A. Tsukanova

**39. Approaches to the Resource Management in Large Social Networks**

# Supervisor: Olga A. Tsukanova