Утверждено академическим руководителем

Образовательной Программы

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

уровень образования: магистратура

 одобрено Академическим Советом программы

18.10.2018

1. **Self-Service Analytics Using Process Mining Tools**

***Abstract:*** **Self-Service Analytics**is a  form of business intelligence (BI) in which line-of-business professionals are enabled and encouraged to perform queries and generate reports on their own, with nominal IT support. Self-service analytics is often characterized by simple-to-use BI tools with basic analytic capabilities and an underlying data model that has been simplified or scaled down for ease of understanding and straightforward data access (Gartner IT-Glossary).

**Process mining** aims to discover, monitor and improve real processes by

extracting knowledge from event logsreadily available in today’s information

systems (van der Aalst, W.M.P.: Process Mining: Discovery Conformance and Enhancement

of Business Processes)

The goal is to find benefits from the integration of those two concepts.

References:

1. Process Mining: Data Science in Action by W.M.P. van der Aalst, Springer Verlag, 2016 (ISBN 978-3-662-49850-7).
2. Process Mining: Discovery, Conformance and Enhancement of Business Processes by W.M.P. van der Aalst, Springer Verlag, 2011 (ISBN 978-3-642-19344-6).
3. SAP Process Mining by Celonis <https://www.sap.com/products/process-mining.html>

# Supervisor: Svetlana V. Maltseva

1. **Agility and virality in modern business architecture**

***Abstract:*** Agility is one of the modern properties to be successful business architecture in response to quick market change. It describes possibilities of 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 MSQRD, PRISM and many others demonstrate how important is the integration between information systems, business demands and staff 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. [Zykov](http://jonahberger.com/) S., Gromoff A., Kazantzev N. Software engineering for enterprise system agility, IGI Global, Chapter 3,5, USA, 2019
2. <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>
3. [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](https://www.amazon.com/Invisible-Influence-Hidden-Forces-Behavior/dp/1476759693/ref%3Das_li_ss_tl?ie=UTF8&linkCode=sl1&tag=jbsite-20&linkId=11406899823cffd33264b6d0ea568c6f)
4. <https://www.amazon.com/Contagious-Things-Catch-Jonah-Berger/dp/1451686579/>

# Supervisor: A.Gromoff

1. **Subject oriented approach to (big) Data-intensive workflows in aerospace industry (running project)**

***Abstract:*** Subject-oriented business process management  (S-BPM) refers to a formal notation system for describing and executing business processes. In this system, the focus is on the subject or individual actor. S-BPM differs from other modelling languages in its low number of modelling symbols and close approximation of natural language—and therefore its similarity to the way people generally gather information, think and communicate. With this focus on the subject, actors can model their processes from a first-person perspective and experience them immediately. Because S-BPM processes can be executed immediately after modelling.

In this work a student is expected to find intersections between topics of S-BPM and Big Data. Would processing Big Data influence models built around employees and customers ? Student is expected to draw a set of S-BPM models for the cases of Big Data application in the aerospace industry.

References:

1.     Fleischmann, Albert, et al. *Subject-oriented business process management*. Springer Publishing Company, Incorporated, 2014.

2.     <https://www.metasonic.de/en/s-bpm>

3.     Schiefer, Josef, et al. "Process information factory: a data management approach for enhancing business process intelligence." *e-Commerce Technology, 2004. CEC 2004. Proceedings. IEEE International Conference on*. IEEE, 2004.

4.     Fleischmann, Albert, et al. "Subject-oriented modeling and execution of multi-agent business processes." Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02. IEEE Computer Society, 2013.

**Supervisor: A.Gromoff**

1. **Social reflection in modern trends of Digital Transformation**

***Abstract:*** In today literature many aspects of social mass, economy and modern technologic interrelations were studied and observed, but several aspects such as readaptation of elder part or reeducation of younger workless part and others concerns are either weakly touched or even skipped from observation and consideration. As well problem of semantic unevenness in communication between different social classes is marginal. Thus, the work is dedicated to consider mentioned problem and aspects of modern life.

In this work a student is expected to provide wide scale literature overview and to find intersections between mentioned topics and development of the initiatives Industry 4.0, Digital Transformation and Digital Economy.

 References:

1. [Zykov](http://jonahberger.com/) S., Gromoff A., Kazantzev N. Software engineering for enterprise system agility, IGI Global, Chapers 3,4,5, USA, 2019
2. Lowall A., Schalter T., Reicheit D., Role and Rights Management, in S-BPM in the Wild (edts Fleishman S., Schmidt W., Stary Ch.), Springer open 2015

**Supervisor: A.Gromoff**

1. **An approach to interlaying modeling of business process in administrational governance.**

Abstracts: Modern methods of BP modelling are mainly overwhelmed with broad and often useless notation, as long some can be too laconic and abstract for mere mortal users. In the work an attempt to join best practices from subject oriented use and from graffiti icon construction is presented to simplify the process of modelling and adaptation to process management in the circle of employee of administrational governance, were lack of time for reeducation and well-established rules/habits are main obstacles for effective implementation of process management principals.

In this work a student is expected to provide an attempt to develop new method of process description and thus modeling on the platforms of S-BPM and macro-joined pictograms.

Reference:

* 1. Fleischmann, A., et al. *Subject-oriented business process management*. Springer Publishing Company, Incorporated, 2014.
	2. Becker J., Pfeiffer D., Rackers M. Domain specific process modelling in public administration – The PICTURE approach, Electronic government 6th international conference, EGOV 2007, Regensburg (eds. Welmmer M., Scholl J.)

**Supervisor: A.Gromoff**

1. **Big data Analysis and Industry 4.0 / Smart Manufacturing**

***Abstract:*** Industry 4.0 represents the backbone of 4th industrial revolution that is envisioned to transform conventional manufacturing and supply chains into networks of interconnected cyber-physical systems (smart factories) that reduces the dependence on human labour and outsourcing necessity. The benefits of such solutions comprise the raise of productivity, enforced robotization, supply chain transparency and sustainability, elimination of child labour, real-time quality assurance and mass customization.

Industry 4.0 is driven by interconnectivity due data sensors that connect production lines and allow injecting customer preferences directly into the product assembly. However, the collections of data from sensors are left unanalysed caused by the lack of industrial evidence, cost-benefit analysis and reported gains.  Big Data Systems in the Industry 4.0 solutions might provide new insights and improve the efficiency of decision making that represents a research gap in this area.

The task for student here would be:

1) To make an overview of academia and industrial sources about Industry 4.0, Big Data and the intersection of both

2) To select the gap that would define his/her unique research stream and the applicable research methods together with the supervisor

3) To work in a tight collaboration with a supervisor (ideally a sprint = 1week) to execute the study

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.  Hermann, M., Pentek, T. and Otto, B., 2016, January. Design principles for industrie 4.0 scenarios. In System Sciences (HICSS), 2016 49th Hawaii International Conference on (pp. 3928-3937). IEEE.

2. Kagermann, H., 2015. Change through digitization—Value creation in the age of Industry 4.0. In Management of permanent change (pp. 23-45). Springer Gabler, Wiesbaden.

3. Anna, K. and Nikolay, K., 2015. Survey on big data analytics in public sector of russian federation. Procedia Computer Science, 55, pp.905-911.

4. Fay M., and Kazantsev N., 2018. When Smart gets Smarter: How Big Data Analytics Creates Business Value in Smart Manufacturing. ICIS 2018 Proceedings

**Supervisor N.Kazantsev**

1. **Big data Analysis and Blockchain**

***Abstract:*** Blockchain represents the emerging area of research comprising various protocols of constructing decentralised ledger in industries. Whilst being implemented, such solutions lack transaction speed and new ways to increase its efficiency. In particular, (big) data generated in industries might be analysed and applied for improving certain characteristics of decentralised ledger solutions.

The task for student here would be :

1) to select one or several industries (e.g. aerospace, automotive, telecommunications...  ) where Block chain solution is (planned to be) implemented

2) to make an overview of academia and industrial sources about Blockchain, Big Data and the intersection of both

3) To select the gap that would define his/her unique research stream and the applicable research methods together with the supervisor

4) To work in a tight collaboration with a supervisor (ideally a sprint = 1week) to execute the study

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. Mendling, J., Weber, I., Aalst, W.V.D., Brocke, J.V., Cabanillas, C., Daniel, F., Debois, S., Ciccio, C.D., Dumas, M., Dustdar, S. and Gal, A., 2018. Blockchains for business process management-challenges and opportunities. ACM Transactions on Management Information Systems (TMIS), 9(1), p.4.

**Supervisor N.Kazantsev**

1. **Ethics of Big data Analysis**

***Abstract:*** The mass amount of individual data collected and operated by individuals raises concerns of policy-makers. The recent GDPR policy represents the increasing awareness that European Union has about Big Data Analytics in companies. The recent cases about potential election interference in the US, Huawei spying scandal and many others recalls for ethical consideration of data classification, information privacy and sustainable data management.

The task for student here would be :

1) to select one or several industries (e.g. aerospace, automotive, telecommunications...  ) where Big data is collected

2) to make an overview of academia and industrial sources about Information Privacy, Big Data Ethics and the intersection of both

3) To select the gap that would define his/her unique research stream and the applicable research methods together with the supervisor

4) To work in a tight collaboration with a supervisor (ideally a sprint = 1week) to execute the study

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. Cadwalladr and Graham-Harrison, 2018 “Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach” URL: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.

2. Giber, L. (2016). Ethical framework of big data application. In *Proceedings of International Conference Information Systems 2016* (pp. 1-6). URL: <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1010&context=sigbd2016>

**Supervisor N. Kazantsev**

1. **Big data Systems in the selected industry/company**

***Abstract:*** The application of Big Data Analytics has crossed almost all industrial borders and is being adopted by the major large-scale companies and governments. Currently, there is a scarcity on research on industrial differences of Big data application, i.e. How these kind of systems are being installed, how they are utilised, what are the strengths and weaknesses, threats and opportunities, benefits and barriers that exist whilst their life-cycle.

The task for student here would be:

1) to select one or several industries (e.g. aerospace, automotive, telecommunications...  ) where Big data is collected

2) to make an overview of academia and industrial sources about Big Data Application in different industries

3) To select the gap that would define his/her unique research stream and the applicable research methods together with the supervisor

4) To work in a tight collaboration with a supervisor (ideally a sprint = 1week) to execute the study

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. Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275, 314-347.
2. Proceedings Special Interest Group of Big Data Application at the International Association for Information Systems (AIS) <https://aisel.aisnet.org/sigbd/>

**Supervisor N. Kazantsev**

1. **COGNITIVE AND INTELLIGENCE DECISION MAKING TOOLS SUPPORTING GOVERNENCE AND STRATEGIC CONTROL OF COMPLEX SYSTEM DEVELOPMENT**

***Abstract:*** Modern decision maker’s situation is characterized not only by rapid variability, by the interaction of many diverse and interdisciplinary factors, but also by the presence of active stakeholders whose beliefs and interests lead to variety of the directions of the situation development. The decision making on governance of complex socio-economic system and situation (socio-political, cultural, economical and other ) mostly depends on collective decision making, experts experience and available information analysis. In the field of enterprise governance or public administration including the formation of strategy or public policy (government policy making), governance in condition of strategic development becomes increasingly important along with traditional management. Management is the attainment of organizational goals in an effective and efficient manner through planning, organizing, leading, and controlling organizational resources. Governance denotes other activities, mainly related to goal setting and monitoring (monitoring). Governance ensures that stakeholder needs, conditions and options are evaluated to determine balanced, agreed-on system objectives (goals) to be achieved; setting direction through prioritization and decision making; and monitoring performance and compliance against agreed-on direction and objectives.

There are varieties of challenges for solve typical task in governance cycle(goal-setting based on external environment prediction and positioning in space of stockholders interest , 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) in condition of complex rapidly changed situations:

* 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).
* Uncertainty that call by rapid condition changes, variety of possible scenarios of future development. Uncertainty of development goals of ill-structured system and criteria for choosing control solution can also be noted. As a rule, dissatisfaction with current condition of system is realized by an individual, but his knowledge of causes and possible means of changing situation in ill-structured system are fuzzy and conflicting. Formalization of fuzzy representations is one of the main problems that have to be solved while developing models and methods of making decisions in ill-structured situations.
* Another difficulty consists in that subjects of governance have to manipulate qualitative information in form of hypotheses (assumptions), intuitive concepts, and semantic images. Numerous studies of decision making processes confirm that the subject of control is unusual for thinking and making decisions only in quantitative terms. He thinks mainly qualitatively and sees the solution finding process as, first of all, searching of solution idea, where quantitative estimations play auxiliary role.
* Lack of time for the working-out decision based on deep analysis It is also important to take into account that the subject of control very often has to make decisions in permanently varying conditions and limited endurance

Despite on rapidly growth of data volume describing system and processes and technological and analytical tools for its analysis, listed challenges have not lost their relevance and just have been transformed. 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 state or 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. Existing decision support systems focus on amateurs but decision maker and problem solver are not supported by modern system. This leads to resistance to implement of analytical tools and made decisions do not taking to account results obtained in analytical system. Discrepancy between decision makers, problem solvers and confidence level of experts to results of analytical information systems will grow. Therefore, the level of maturity of information system doesn’t support decision making in holistic cycle of development governance.

This situation has resulted in necessity of creation of integrated systems of support of goal setting process and making of administrative decisions when working out a strategy of system development. Decision support systems supported development governance should be based:

* on cognitive approach used (i) to support collective deep insight (in opposite deep search) based semantic recognizing under influence experts model of situation representation and (ii) to embedded possibility of cognitive control in process governance problem solving;
* self-organization principals used (i) to consider the governance processes in space of interrelation of stakeholders interest and then (ii) to provide assembling of decision making tools involving in governance processes by active stakeholders’ through the use of new information and communication technologies (Internet of Things (and everything), Blockchain, Big data tools and other ).

Considering the foregoing, the following topics are suggested for students' studies in course and diploma theses:

1. Tools for strategy correction in condition of changing business environment
2. Week signal revealing based on big data tools.
3. Analysis of complex system development based on fuzzy-cognitive mapping

**Supervisor: Zinaida K. Avdeeva**

1. **TOOLS FOR STRATEGY CORRECTION IN CONDITION OF CHANGING BUSINESS ENVIRONMENT/**Инструменты коррекции стратегии развития системы в режиме отслеживания изменений во внешней среде

**Abstract:** The study raleted with problem of strategy monitoring and following correction of strategy because of exernal environment change and lead to problem of goals achivment. It need to develop approach to strategy control based on revealing the problem situation (considered as deviation between goals state and achievable state in existing condition) and solving it in whole cycle of external and internal changes monitoring.

In general, the control of system can be represented as construction of strategy for the system development, defining the main goals and general directions for their reaching, and its implementation. Revealing the system development problems influencing negatively achievement of strategic control goals is one of the key stages of construction of strategy.

Thus, the control problem consists in transfer of system into one of a nearest state corresponding to goal image. At that, the proposed approach allows determining the system state in both values of model factors and rates of factor changes. We presents the general scheme of method for strategy monitoring on the base of linear dynamic models on the base of cognitive map in regard to business systems .

Strategic monitoring plays one of the key roles in a cycle of supporting the formation, implementation and correction of the socio-economic system (SES) (industrial enterprises, corporations, cities, etc.) development. Its main purpose is to monitor the achievement of strategic goals of a system development in a changing external environment. Strategic monitoring is aimed at the timely detection of (1) favorable and unfavorable changes in the external environment, and (2) changes in the system properties (its strengths and weaknesses), which may affect the implementation of the adopted development strategy.

Considering complexity, weak formalizability of modern situations, on the one hand, strategic management experts note the necessity of expansion of the corresponding scientific and methodical support. On the other hand, methods of analysis and forecasting of system development on the basis of casual models, have proven to be pretty effective. Casual model (for example, cognitive map of a situation) is a model representing the expert knowledge of a situation in the form of structure of causal influences.

References:

* Avdeeva Z., Kovriga S. The technology of the strategic goal-setting and monitoring of a system development on the basis of cognitive mapping In: Procedia Computer Science. 4rd International Conference on Information Technology and Quantitative Management, ITQM 2017  - IN PRINT.
* 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.
* 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.
* 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.
* Kahraman, Cengiz et al. (2008) Fuzzy Multi-Criteria Decision Making -Springer. Pp.380
* [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**

1. **WEEK SIGNAL REVEALING BASED ON BIG DATA TOOLS/**Выявление слабых сигналов изменения ситуаций на основе инструментов работы с большими данными

**Abstract:** For the problem solving there are hard and soft approaches to system modeling, monitoring. At that, ill-conditioned systems are characterized by problems that hardly can be extracted from analyzed control situation. This limits applicability of traditional methods for finding optimal (or even satisfactory) solution to control problems for such systems. In this process, we consider the intelligent activity of problem solver who should deep insight in problem situation of system under control and could offers a strategic decision based on analysis different scenarios. Problem solver is considered as a group expert thus we have to analysis and monitor a lot of information source: expert opinions and big data sets.

The characteristic feature of ill-structured system analysis consists in that the process of preparation and making decisions on control of ill-structured system is, as a rule, a group activity. Each participant of this process represents problem situation based on “his own” inherent representations and knowledge of situation (images, models of world). An image of world includes a set of convictions, perceptual features, cost and practical rules of an individual that guide his activity and influence the process of problem situation resolution. Thus, decision preparation and making in control problems for ill-structured systems should be considered as complex intellectual process of problem resolution that cannot be reduced to solely rational choice. 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.

* Information overload and information lack exist simultaneously.
* Basic principal of big data approach to statistical data analysis leads to loss possibility of weak signal analysis. In case of complex system development except for typical scenarios unexpected can be actualized. From the point of view supporting IT-system it’s needed only monitor of changes of putted factors of expert’s scenario. Our practical experience shows disadvantages of data-mining tools without supervising model, which can only reveal statistically significant influences, factors, or data clusters in general.
* The significance problem for analytical decision making is problem of data quality. For the data cleansing spend huge resources, but it is fact that more than 40% from total volume of data is “dirty”. There are variety of industrial tools to data cleansing from syntaxes mistakes and mistakes with missing data, but it remain time-consuming. There are two major reason of dirtiness of data: human induced (directly or indirectly embedded in IS especially in interface for data input); machine induced (technical interrupt leading to appear mistakes on a stage of data saving, updating.

It is important to note that the observed factors of the situation are heterogeneous, and accordingly, they have different arrays of accumulated data and information that can not use similar monitoring tools.

In the present conditions of increasing instability and the speed of changes in situations, it is increasingly difficult to predict the nature of these changes and respond in a timely manner to them. Therefore, the development of approaches to monitoring as systematic monitoring of the system parameters with the purpose of revealing weak signals - early and odd signs of the onset of crisis situations, events that are essential for the development of the system - is becoming more important. (The concepts of weak and strong signals, as well as related levels of awareness of imminent changes, were introduced by I. Ansoff [5].)

In the framework of studies conducted to monitor the effect of weak effects of changes in observed factors.

References:

1. Avdeeva Z., Raikov A., Ermakov A. [Big Data Refining on the Base of Cognitive Modeling](https://publications.hse.ru/view/198195185) // IFAC-PapersOnLine. 2016. Vol. 49. No. 32. P. 147-152. [doi](http://doi.org/10.1016/j.ifacol.2016.12.205%22%20%5Ct%20%22_blank)
2. Ansoff, I. H. Implanting strategic management. New Jersey: Prentice Hall. 1984.
3. Ansoff, I. H. Managing strategic surprise by response to weak signals. California Management Review, 18(2), 21-33. 1975.
4. Ansoff, Igor. Implanting Strategic Management. Prentice // Hall International Inc. 1984.
5. Cooper A. Weak Signals and Text Mining I – An Introduction to Weak Signals. URL: http://blogs.cetis.org.uk/adam/2011/05/12/weak-signals-andtext-mining-i-an-introduction-to-weak-signals/
6. Christos Ch., Awais R. and Paul J. Taylor. Weak Signals as Predictors of Real-World Phenomena in Social Media // 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). Pp. 864-871. DOI: 10.1145/2808797.2809332
7. Ponomareva Julia V., Sokolova Anna V.. The identification of weak signals and wild cards in foresight methodology: stages and methods // Working papers. Series: Science, Technology and Innovation. National Research University Higher School of Economics. WP BRP 46/STI/2015. 2015. 26 p.
8. Tabatabaei N.. Detecting Weak Signals by Internet-Based Environmental Scanning // Waterloo, Ontario, Canada, 2011.
9. Yoon, J. Detecting weak signals for long-term business opportunities using text mining of Web news // Expert Systems with Applications, 39(16), 2012. 12543-12550.

**Supervisor: Zinaida K. Avdeeva**

1. **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
6. Avdeeva Z., Raikov A., Ermakov A. [Big Data Refining on the Base of Cognitive Modeling](https://publications.hse.ru/view/198195185) // IFAC-PapersOnLine. 2016. Vol. 49. No. 32. P. 147-152. [doi](http://doi.org/10.1016/j.ifacol.2016.12.205%22%20%5Ct%20%22_blank)

**Supervisor: Zinaida K. Avdeeva**

**14. The Creation Administrative and Automation Tools for Database of Big Data Project on Base Radioastronomy Data/Создание административных инструментов и средств автоматизации базы данных проекта больших радиоастрономических данных.**

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++.

Scientific adviser: Dr. Samodurov V.A.1,2 (<http://www.hse.ru/org/persons/37253680>),

1 National research university Higher school of economics, Moscow, Russia

2Pushchino Radio Astronomy Observatory ASC LPI, Pushchino, Russia

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: Samodurov V.A**

**15. The processing of large radio astronomical data using a GPU (NVIDIA graphics cards)/Обработка больших радиоастрономических данных при помощи GPU (графических плат NVIDIA)**

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 instrument for stream computing of these big data. We need in launch the streaming data on various types of high-performance computing systems, including using a GPU (NVIDIA graphics cards) [4]. You must start-up, configuring and run NVIDIA graphics cards by CUDA, translate in CUDA language of some computing and astronomical algorithms (sorting, correlation, Fourier analytics etc. also as some astronomical algorithms). Need skills from You: C++ , CUDA technology, SQL (Postgresql preferable [5]), good skills also will Perl or/and Python.

Scientific adviser: Dr. Samodurov V.A.1,2 (<http://www.hse.ru/org/persons/37253680>),

1 National research university Higher school of economics, Moscow, Russia

2Pushchino Radio Astronomy Observatory ASC LPI, Pushchino, Russia

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://www.nvidia.ru/object/cuda-parallel-computing-ru.html>
5. <https://www.postgresql.org/docs/9.5/static/docguide.html>

**Supervisor: Samodurov V.A**

**16. Processing of large radio astronomical data using supercomputer computations/Обработка больших радиоастрономических данных при помощи суперкомпьютерных вычислений**

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 instrument for stream computing of these big data. We need in launch the streaming data on various types of high-performance computing systems, including using supercomputer computations. [4]. You need must translate on supercomputer of some computing and astronomical algorithms (sorting, correlation, Fourier analytics etc. also as some astronomical algorithms). Need skills from You: C++, high performance computing technology, SQL (Postgresql preferable [5]), good skills also will Perl or/and Python.

Scientific adviser: Dr. Samodurov V.A.1,2 (<http://www.hse.ru/org/persons/37253680>),

1 National research university Higher school of economics, Moscow, Russia

2Pushchino Radio Astronomy Observatory ASC LPI, Pushchino, Russia

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://parallel.ru/>
5. <https://www.postgresql.org/docs/9.5/static/docguide.html>

**Supervisor: Samodurov V.A**

**17. Analysis of the approaches to designing the systems of cross-lingual intelligent information access/Анализ подходов к разработке систем межъязыкового интеллектуального доступа к информации.**

Abstract: During last decade, one has been able to observe a quickly growing interest in the design of computer intelligent agents fulfilling cross-lingual information retrieval (CLIR) on the Web. It is a consequence of emerging a huge, permanently increasing number of Web-sources in languages being different from English. In September 2012, a seminar on Multilingual Semantic Web (MSW) was organized in Germany in the Dagstuhl Castle. The proceedings of this seminar contain the following data: in the year 2010, the number of non-English-speaking Internet users was three times higher than the number of English-speaking users (1430 million vs. 536 million users). That is why the problem of developing a MSW is very topical. It is broadly accepted that a promising approach to the realization of CLIR on the Web is employing a special semantic language-intermediary (SLI) in order to represent in the same format both semantic content of a user query and semantic content of the analysed fragment of a text in natural language (NL). The task of the term paper is to analyse the principoal existing approaches to designing the systems of cross-lingual intelligent information access, using a semantic language-intermediary.

References:

 1. Fomichov, V. A. [SK-languages as a Powerful and Flexible Semantic Formalism for the Systems of Cross-Lingual Intelligent Information Access](https://publications.hse.ru/view/208921070) // Informatica. An International Journal of Computing and Informatics. 2017. Vol. 41. No. 2. P. 221-232.

2. 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.

3. 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.

4. 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.

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).

7. 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)

8. 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

**18.The analysis of knowledge-based approaches to natural language processing in business informatics/Анализ основанных на знаниях подходов к компьютерной обработке естественного языка в бизнес-информатике**

Abstract: Business Intelligence (BI) is a term used for denoting the models, methods, and applied computer systems that aim to provide businesses with support for decision-making  by means of collecting and processing information relevant to business leaders. The huge amount of textual information that is available in online sources and in corporative data bases makes it nearly impossible to follow and analyse all the relevant sources manually or to apply traditional storage and retrieval methods. Useful business information is hidden in  the mass of data that is permanently increasing and evolving.

Information Extraction (IE) deals with extracting relevant and high-quality information  from unstructured sources, , such as texts in natural language (NL) and document collections. The extraction process involves representing information contained in textual data  in a structured and normalized way. The Ontology-Based Information Extraction (OBIE) combines the processes of extracting and storing information and uses an ontology (a knowledge base) for extracting significant business events from NL-texts. The most broad prospects for representing the extracted information in a formal way are opened the Theory of K-representations (knowledge representations) developed by V. A. Fomichov. This theory includes, in particular, a new class of formal languages called SK-languages (standard knowledge languages).

References

1.      Arendarenko, E., Kakkonen, T. Ontology-based information and event extraction for business intelligence. In: A. Ramsay. G. Agre (Eds.). Artificial Intelligence: Methodology, Systems, and Applications. 15th Intern. Conference, AIMSA 2012, Varna, Bulgaria, September 2012, Proceedings. Springer, Lecture Notes in Artificial Intelligence, 2012, Vol. 7557, p.89-102.

2.      Wang C., Xiong M., Zhou Q., Yu Y. PANTO: a portable NL-interface to ontologies 4th European Semantic Web Conference Proceedings. Springer, 2007, p. 473-487.

3.      Fomichov, V.A.: Semantics-Oriented Natural Language Processing: Mathematical Models and Algorithms. Springer, New York, Dordrecht, Heidelberg, 2010.

4.      Razorenov, A. A., Fomichov, V. A. [A new formal approach to semantic parsing of instructions and to file manager design](https://publications.hse.ru/view/194073739), In: Database and Expert Systems Applications. 27th International Conference, DEXA 2016, Porto, Portugal, September 5-8, 2016, Proceedings Vol.. 9827. Part I. London, New York, Dordrecht, Heidlberg, Cham: Springer, 2016. P. 416-430.

# Supervisor: Vladimir A. Fomichov

**19. Большие данные в биоинформатике и здравоохранении / Big Data in Bioinformatics and Healthcare**

Abstract: Big data technologies are increasingly used for bioinformatics and health-care research. Large amounts of biological and clinical data have been generated and collected at an unprecedented speed and scale. Progress in bioinformatics and healthcare can be achieved only through effective, correct use of tools for big data processing. The role of big data techniques in bioinformatics applications is to provide data repositories, computing infrastructure, and efficient data manipulation tools for investigators to gather and analyze biological information. Biomedical scientists are facing new challenges of storing, managing, and analyzing massive amounts of datasets. For example, DNA sequencing is on the path to becoming an everyday tool in medicine; the new generation of sequencing technologies enables the processing of billions of DNA sequence data per day. However, there lack efficient big data architectures and tools for many important bioinformatics problems.

Term paper can be devoted to the following issues: Big data problems in bioinformatics; Methodologies and tools for big data analytics in bioinformatics; Computational biology perspective in big data era: challenges and future goals; Advanced datamining using RNAseq data; Analysis of genomic data in a cloud-computing environment; etc.

References

1. V. Marx, “Biology: The big challenges of big data,” Nature, vol. 498, no. 7453, pp. 255–260, 2013.
2. B. Wang, R. Li, W. Perrizo, “Big Data Analytics in Bioinformatics and Healthcare,” IGI Global (USA), 2014, 528 p.
3. M.D. Lytras, P. Papadopoulou, “Applying Big Data Analytics in Bioinformatics and Medicine,” IGI Global (USA), 2017, 465 p.
4. “Bioinformatics: Concepts, Methodologies, Tools, and Applications,” Ed.: Information Resources Management Association (USA), IGI Global (USA), 2013, (3 Volumes) 1783 p.
5. M. Reshma, M.A. Supaksha, N. Hemalatha “Application of Big Data in Bioinformatics,” Int. J. of Latest Trends in Engineering and Technology, Special Issue SACAIM 2016, pp. 206-212.
6. Jake Luo, Min Wu, Deepika Gopukumar, Yiqing Zhao “Big Data Application in Biomedical Research and Health Care: A Literature Review,” Biomed Inform Insights. 2016; 8: 1–10.

# Supervisor: Vasily V. Kornilov

**20. Большие данные в персонализированной медицине / Big Data in Personalized Medicine**

Abstract: Healthcare Informatics combines information and understanding from medical areas (pre-clinical, clinical and post-clinical), healthcare administration and management and information technology. Big data applications present new opportunities to discover new knowledge and create novel methods to improve the quality of health care. The application of big data in health care is a fast-growing field, with many new discoveries and methodologies to make people better. The large volumes of data from bioinformatics and healthcare informatics domains coupled with analytics can deliver preventive, predictive and personalized healthcare aids.

Term paper can be devoted to the following issues: The benefits of big data analytics in the healthcare sector; Methodologies and tools to manage massive and complex information of electronic medical records; Personalized disease phenotypes from massive OMICs data

References

1. C.E. Morr “Research Perspectives on the Role of Informatics in Health Policy and Management,” IGI Global (USA), 2013, 323 p.
2. R. Wachter “The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine’s Computer Age,” McGraw-Hill Education, 2017, 352 p.
3. L.L. Liang “Connected for Health: Using Electronic Health Records to Transform Care Delivery,” Pfeiffer, 2017, 272 p.
4. A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D.A. Beard, K. Najarian “Big Data Analytics in Healthcare,” BioMed Research International, Volume 2015, Article ID 370194, 16 pages
5. S. S.-L. Tan, G. Gao, S. Koch “Big Data and Analytics in Healthcare,” Methods Inf Med 2015; 54: 546–547

# Supervisor: Vasily V. Kornilov

**21. Цифровое преобразование бизнеса / Digital Business Transformation**

Abstract: “Digital business is the creation of new business designs that not only connect people and business, but also connect people and business with things to drive revenue and efficiency. Digital business helps to eliminate barriers that now exist among industry segments while creating new value chains and opportunities that traditional businesses cannot offer.” (Gartner, 2015).

Digital Transformation radically alters markets through the application of new digital technologies, and it challenges established business models. Never before has IT played a more pivotal role. In the age of IoE (the Internet of Everything) and Big Data, IT has become the business process. IT is helping companies develop new business models to adapt to market and competitive forces, fundamentally changing companies’ relationships with their customers. Digital business is built on new computing infrastructure – the pillars of mobile, cloud, Big Data, and analytics – accelerated by the Internet of Things (IoT), advances in machine learning, and innovations like blockchain. These disruptive technologies are giving companies the ability to radically change business models, and create new products and services.

The future of every industry is digital. Fully digital business processes will be intelligent systems that respond to any situation, connecting with both machines and people to enable critical, real-time responses and insight-driven decision-making.

What is driving digital transformation? What a digital strategy is? How do organizations embark upon digital business transformation? How the digital strategy can serve business objectives?

References

1. Martin Gill, Ted Schadler, “The Digital Business Transformation Playbook For 2017,” Forrester Research, 2017 // https://www.forrester.com/report/Digitize+Your+Business+Strategy/-/E-RES115755
2. Bernard Marr “IT Playing Strategic Role In Digital Business Transformation”, PARS International, 2017 // https://www.forbes.com/sites/bernardmarr/2017/08/21/it-playing-strategic-role-in-digital-business-transformation
3. David Rogers “The Digital Transformation Playbook: Rethink Your Business for the Digital Age”, Columbia University Press, 2016, 296 p.
4. Venkat Venkatraman “The Digital Matrix: New Rules for Business Transformation Through Technology”, LifeTree Media, 2017, 224 p.
5. Isaac Sacolick “Driving Digital: The Leader's Guide to Business Transformation Through Technology”, AMACOM, 2017, 224 p.

# Supervisor: Vasily V. Kornilov

**22. Domain adaptation using neural networks /Использование нейронных сетей для адаптации в обучении**

Abstract: Practical machine learning often involves handling small datasets. This restricts usage of powerful models such as neural networks. Although large datasets (that are somehow similar) are often available and can help solving the original problem. For example, one can use classification model trained on Wikipedia pages to classify small enterprise documents. It is known that neural networks are well fitted for such problems. This work includes learning essentials of neural networks; research on domain adaptation with neural networks; empirical comparison of these approaches on available datasets.

*Yosinski, Jason, Jeff Clune, Yoshua Bengio, and Hod Lipson. “How transferable are features in deep neural networks?.”* [*http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks*](http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks)

# Supervisor: Sergey Lisitsyn

**23. Multi-task learning using neural networks /Использование нейронных сетей для многозадачного обучения**

Abstract: One of promising subfields of machine learning is multitask learning. This way of learning is based on solving multiple similar problems at the same time. The approach is often helpful as different models may ‘help’ each other. For example, it is known that a person that speaks a few languages learns new languages faster. Architectural flexibility of neural networks fits multitask learning quite well. This work includes learning essentials of neural networks and modern approaches for multitask learning. Practical part of the work includes empirical research of multitask learning techniques using open datasets.

*Sebastian Ruder. “An Overview of Multi-Task Learning in Deep Neural Networks”,* [*https://arxiv.org/pdf/1706.05098.pdf*](https://arxiv.org/pdf/1706.05098.pdf)

# Supervisor: Sergey Lisitsyn

**24. Comparison of fastText and word2vec for text classification / Сравнение fastText и word2vec для задач классификации текстов**

Abstract: Lately, machine learning community has shifted to vector models for textual data. This shift started when word2vec was first introduced. The word2vec algorithm enabled us to build effective semantic vector models of natural languages. The fastText algorithm is a further improvement which also considers syntax similarities of words. This work includes learning basics of vector models for natural languages and empirical comparison of the methods using available datasets.

*FastText and Gensim word embeddings* [*https://rare-technologies.com/fasttext-and-gensim-word-embeddings/*](https://rare-technologies.com/fasttext-and-gensim-word-embeddings/)

# Supervisor: Sergey Lisitsyn

**25. Real time signal (time series) processing/Обработка сигналов (временных рядов) в реальном времени.**

Abstract: The goal of this project is to implement and analyze optimal time series processing with a sliding window.

Main phases:

1. Generation of synthetic data: random function with specified stochastic properties and its measurement.
2. Computation of the optimal influence function of the processing sliding window with the given width and delay.
3. Analysis of the relation between processing quality, delay and the width of the sliding window. Such analysis should allow choosing a good compromise between the reconstruction quality, delay, and computational complexity.
4. Illustrations of the processing performance under different settings.

References:

1. Golubtsov P.V., Starikova O.V. Reduction of parametrically determined invariant measuring systems // Moscow University Physics Bulletin, V 56, No 6, с. 1-4, 2001
2. Filatova S. A., Golubtsov P. V. Invariant Measurement Computer Systems // Pattern Recognition and Image Analysis, 1, N 2, P. 224–235, 1991.
3. Filatova S. A., Golubtsov P. V. Invariance Considerations in Design of Image Formation Measurement Computer Systems // Proceedings of SPIE, V. 1960, Automatic Object Recognition III, P. 483–494, 1993.
4. Голубцов П. В., Филатова С. А. Оптимальная локальная редукция для пространственно-инвариантных измерительных систем // Математическое моделирование, 2, N 11, С. 61–66, 1990.

Supervisor: Peter V. Golubtsov

**26. Real time image processing for potentially infinite field of view /Обработка изображения с потенциально бесконечным полем рения.**

Abstract: The goal of this project is to implement and analyze optimal image processing with a sliding window.

Main phases:

1. Generation of synthetic data: measurement of an image, which includes smoothing-like distortion and random noise.
2. Computation of the optimal point spread function of the processing sliding window with the given size.
3. Analysis of the relation between processing quality and the size of the sliding window. Such analysis should allow choosing a good compromise between the reconstruction quality and computational complexity.
4. Illustrations of the processing performance under different settings.

References:

1. Filatova S. A., Golubtsov P. V. Invariant Measurement Computer Systems // Pattern Recognition and Image Analysis, 1, N 2, P. 224–235, 1991.
2. Filatova S. A., Golubtsov P. V. Invariance and Synthesis of Optimum Image Formation Measurement Computer Systems // Artificial Intelligence, Expert Systems and Symbolic Computing, Elsevier Science Publishers B. V. (North-Holland), P. 243–252, 1992.
3. Filatova S. A., Golubtsov P. V. Invariance Considerations in Design of Image Formation Measurement Computer Systems // Proceedings of SPIE, V. 1960, Automatic Object Recognition III, P. 483–494, 1993.
4. Golubtsov P.V., Sizarev D.V., Starikova О.V. Synthesis of optimal invariant image generation systems on a plane // Moscow University Physics Bulletin, V 58, No 2, pp 1-5, 2003

**Supervisor: Peter V. Golubtsov**

**27. Calibration in time series processing problems/Калибровка в задачах обработки временных рядов.**

Abstract: The goal of this project is to implement and analyze optimal time series processing for the system with the unknown measurement system (which represents distortion of the time series). It is assumed that originally the transformation, distorting the signal (time series) is initially unknown. As a result the reconstruction of the signal is not possible. However, we are able to obtain the true input signal with a certain delay. Such data can be used for the calibration of the measurement system.

Main phases:

1. Generation of synthetic data: random function with specified stochastic properties and its measurement.
2. Extraction of calibration data and accumulating it in an appropriate canonical form.
3. Gradual recalibration of the measurement system, based on the accumulated calibration canonical information.
4. Computation of the optimal influence function of the processing sliding window for the calibrated measurement system for various stages of calibration.
5. Illustrations of the processing performance under different settings.

References:

1. Filatova S.A., Golubtsov P.V. Invariance and Calibration in Image Processing Problems // The Third International Conference on Expert Systems for Numerical Computing. Computer Science Technical Reports Purdue University, pp 97-100 1993
2. Golubtsov P.V., Starikova O.V. Reduction of parametrically determined invariant measuring systems // Moscow University Physics Bulletin, V 56, No 6, с. 1-4, 2001
3. Голубцов П.В., Старикова О.В. Калибровка инвариантных преобразователей информации // Information Processes (Информационные процессы), том 1, № 1, с. 78-88 2001 [www.jip.ru/2001/golub-1.pdf](http://www.jip.ru/2001/golub-1.pdf)
4. Голубцов П.В., Старикова О.В. Проблема калибровки для инвариантных измерительно-вычислительных систем // Труды II международной конференции «Проблемы управления и моделирования в сложных системах», Издательство РАН Самара, с. 143-148. 2000
5. Голубцов П.В., Старикова О.В. Инвариантность и калибровка измерительно-вычислительных систем в непрерывно-дискретном случае // Труды III международной конференции «Проблемы управления и моделирования в сложных системах», Издательство РАН Самара, с. 635-641. 2001

**Supervisor: Peter V. Golubtsov**

**28. Crowding behavior in a supermarket and sales/Поведение толпы в супермаркете и продажи.**

Abstract: The goal of this project is to develop tools and visualization methods for analysis of customers’ flows in supermarkets and its relations to sales. The study will be based on real data, which includes RFID tracking of shoppers’ trajectories.

Main phases:

1. Visualization and comparison of customers’ flows during peak hours and off-peak hours.
2. Revealing relations between crowding behavior and sales.
3. Analysis of main shopping patterns.
4. Crowd dynamics analysis. Detection of zones with chaotic and regular flow.
5. Experimentation with various visualization techniques, such as: heat maps, vector fields, etc.
6. Possibility of trajectory prediction at different spots for different shopping patterns.

References:

1. Yada, K.: String Analysis Technique for Shopping Path in a Supermarket. Journal of Intelligent Information Systems 36, 385–402 (2011)
2. Kholod M., Nakahara T., Azuma H., and Yada K. The Influence of Shopping Path Length on Purchase Behavior in Grocery Store. Proceedings of KES 2010, Lecture Notes in Computer Science, 6278, pp.273-280, 2010.
3. Larson, J.S., Bradlow, E.T., Fader, P.S.: An Exploratory Look at Supermarket Shopping Paths. International Journal of Research in Marketing 22(4), 395–414 (2005)
4. Hui, S.K., Fader, P.S., Bradlow, E.T.: Path Data in Marketing: An Integrative Framework and Prospectus for Model Building. Marketing Science 28(2), 320–335 (2009)
5. Hui, S.K., Fader, P.S., Bradlow, E.T.: The Travelling Salesman Goes Shopping: The Systematic Deviations of Grocery Paths from TSP Optimality. Marketing Science 28(3), 566–572 (2009)
6. Hui, S.K., Bradlow, E.T., Fader, P.S.: Testing Behavioral Hypotheses using An Integrated Model of Grocery Store Shopping Path and Purchase Behavior. Journal of Consumer Research 36(3), 478–493 (2009)

**Supervisor: Peter V. Golubtsov**

**29. Fuzzy information and fuzzy experiments in big data context/Нечеткая информация и нечеткий эксперимент в контексте больших данных**

Abstract: Sometimes it is more convenient to represent uncertainty in terms of fuzzy sets as opposed to probability distributions. Fuzzy sets representations may better reflect such information as expert opinions.

The goals of this project are to

1. Design formal representation of an “expert assessment” process as a fuzzy measurement.
2. Develop methods for:
	1. updating fuzzy information, based on such measurement;
	2. extracting information from many independent fuzzy measurements and presenting it in a convenient intermediate form;
	3. combining pieces of such information;
	4. and deriving a decision from the accumulated information.
3. Visualize all the transformations of fuzzy information.

References:

1. Zadeh L. A. Fuzzy Sets // Inform. and Control, 8, P. 338–353, 1965.
2. Golubtsov P.V. Theory of Fuzzy Sets as a Theory of Uncertainty and Decision-Making Problems in Fuzzy Experiments // Problems of Information Transmission, V 30, No 3, pp 232-250, 1994
3. Dubois D., Prade H. Fuzzy Sets and Systems: Theory and Applications. New York: Academic Press, 1979.
4. Golubtsov P. V. Fuzzy logical semantics of Bayesian decision making // SPIE Proceedings Vol. 2493. Applications of Fuzzy Logic Technology II, pp 228–239, 1995.

**Supervisor: Peter V. Golubtsov**

1. **Applications of blockchains in mobile systems**

***Abstract:*** Blockchain technology provides, among other purposes such as cryptocurrencies, means to verify integrity of log information from potentially large systems. One example of an important log in mobile communication systems, is the log of phone calls and data usage. This is important because the log relates to the eventual bills sent to the mobile user. In case of roaming, fees may become surprisingly big, although inside European Union the high roaming fees are vanishing. The study would find out how blockchain technology could be used for such purpose in mobile systems. Alternatively, the study could focus on identifying other use cases of blockchains in mobile communications systems.

References:

1. Bonneau, J., Narayanan, A., Miller, A., Clark, J., Kroll, J.A., Felten, E.W.: Mixcoin: anonymity for Bitcoin with accountable mixes. In: Christin, N., Safavi-Naini, R. (eds.) FC 2014. LNCS, vol. 8437, pp. 486–504. Springer, Heidelberg (2014). https://proxylibrary.hse.ru:2117/10.1007/978-3-662-45472-5\_31
2. Zheng, Z., Xie, S., Dai, H., Chen, X., Wang, H.: An overview of blockchain technology: architecture, consensus, and future trends. In: 6th International Congress on Big Data. IEEE (2017)
3. State of blockchain q1 2016: Blockchain funding overtakes bitcoin (2016)

**Supervisor: Alexandr Gorbunov**

1. **Big data analysis on block chain transactions**

***Abstract:*** Analysing Ethereum transactions and finding out the number of different tokens per block. Identifying incentivise transactions and other interesting facts such as congestion in the network by identifying transaction delay and correlating with real life incidents. And investigate the any kind of patterns in the transaction for some particular patters so the we can predict future fate for some tokens.

References:

1. Bonneau, J., Narayanan, A., Miller, A., Clark, J., Kroll, J.A., Felten, E.W.: Mixcoin: anonymity for Bitcoin with accountable mixes. In: Christin, N., Safavi-Naini, R. (eds.) FC 2014. LNCS, vol. 8437, pp. 486–504. Springer, Heidelberg (2014). https://proxylibrary.hse.ru:2117/10.1007/978-3-662-45472-5\_31
2. Zheng, Z., Xie, S., Dai, H., Chen, X., Wang, H.: An overview of blockchain technology: architecture, consensus, and future trends. In: 6th International Congress on Big Data. IEEE (2017)
3. State of blockchain q1 2016: Blockchain funding overtakes bitcoin (2016)

**Supervisor: Alexandr Gorbunov**

1. **Critical data detection**

***Abstract:*** Edge Computing is a new computing paradigm where server resources, ranging from a credit-card size computer to a small data center, are placed closer to data and information generation sources. With edge computing we can significantly decrease the volumes of data that must be moved. However, how to identify the critical data that we do not want to move.

A thesis project could apply known methods of anomaly or critical data detection to a console log data repository of a distributed monitoring system.

References:

1. Salfner, F., Lenk, M., Malek, M.: A survey of online failure prediction methods. ACM Computing Surveys, CSUR (2010)[Google Scholar](https://scholar.google.com/scholar?q=Salfner%2C%20F.%2C%20Lenk%2C%20M.%2C%20Malek%2C%20M.%3A%20A%20survey%20of%20online%20failure%20prediction%20methods.%20ACM%20Computing%20Surveys%2C%20CSUR%20%282010%29)
2. Basseville, M., Nikiforov, I.V.: Detection of abrupt changes: theory and application. Prentice-Hall, Inc., Englewood Cliffs (1993)[Google Scholar](http://scholar.google.com/scholar_lookup?title=Detection%20of%20abrupt%20changes%3A%20theory%20and%20application.&author=M...%20Basseville&author=I.V..%20Nikiforov&publication_year=1993" \t "_blank)
3. Natella, R., Cotroneo, D.: Emulation of transient software faults for dependability assessment: A case study. In: Proceedings of the Eighth European Dependable Computing Conference, EDCC (2010)

**Supervisor: Alexandr Gorbunov**

1. **Dynamic text indexing with minimizers**

***Abstract:*** Minimizers have become a standard technique in text comparison applications. The idea is to select a representative collection of substrings (through looking at their minimum hash values on several fixed hash-functions) as a fingerprint of for each text, such that the comparison of fingerprints gives an estimate on the similarity of the original texts. The topic of the thesis is to extend the technique to a text indexing setting, where the content of the texts can be modified and one needs to update the already found occurrences for some set of patterns to reflect the modifications. This setting is motivated by a genomic read alignment application, where variation in population is the source of modifications to the reference genome.

References:

1. A. Apostolico and Z. Galil. *Combinatorial Algorithms on Words*. Springer-Verlag, New York, 1985.[zbMATH](http://proxylibrary.hse.ru:2400/MATH-item?0564.00027)[Google Scholar](http://scholar.google.com/scholar_lookup?title=Combinatorial%20Algorithms%20on%20Words&author=A..%20Apostolico&author=Z..%20Galil&publication_year=1985)
2. M. Araújo, G. Navarro, and N. Ziviani. Large text searching allowing errors. In *Proc. WSP’97*, pages 2–20. Carleton University Press, 1997.[Google Scholar](https://scholar.google.com/scholar?q=M.%20Ara%C3%BAjo%2C%20G.%20Navarro%2C%20and%20N.%20Ziviani.%20Large%20text%20searching%20allowing%20errors.%20In%20Proc.%20WSP%E2%80%9997%2C%20pages%202%E2%80%9320.%20Carleton%20University%20Press%2C%201997.)
3. R. Baeza-Yates. Text retrieval: Theory and practice. In *12th IFIP World Computer Congress*, volume I, pages 465–476. Elsevier Science, September 1992.[Google Scholar](https://scholar.google.com/scholar?q=R.%20Baeza-Yates.%20Text%20retrieval%3A%20Theory%20and%20practice.%20In%2012th%20IFIP%20World%20Computer%20Congress%2C%20volume%20I%2C%20pages%20465%E2%80%93476.%20Elsevier%20Science%2C%20September%201992.)

**Supervisor: Alexandr Gorbunov**

1. **Fast nearest neighbor search**

***Abstract:*** There are several approaches to speed up (approximate) nearest neighbor queries in large data sets. Generally, they involve an initial stage where an index data structure is constructed. The index can be used to perform queries when new points arrive. The thesis can review and compare various different approaches (tree-based, locality-sensitive
hashing, random projections, ...) and/or experiment with new variantions of the theme.

References:

1. Zezula, P., Amato, G., Dohnal, V., Batko, M., Similarity Search: The Metric Space Approach, Springer, 2006.
2. V. Hyvönen, T. Pitkänen, S. Tasoulis, E. Jääsaari, R. Tuomainen, L. Wang, J. Corander, and T. Roos (2016). Fast nearest neighbor search
through sparse random projections and voting, in Proc. 2016 IEEE International Conference on Big Data (IEEE Big-Data 2016), Washington
DC, Dec. 5–8.

**Supervisor: Alexandr Gorbunov**

1. **Data mining historical textual traditions**

***Abstract:*** Historical textual traditions, including manuscripts, early printed materials and collections of oral traditions, offer a rich source of information for data mining. Various exploratory data analysis methods can be used to summarize and visualize the data. A thesis project could apply known methods to new data collections available through public repositories or ongoing collaborative projects.

References:

J. Tehrani, Q. Nguyen, and T. Roos, (2016). Oral fairy tale or literary fake? Investigating the origins of Little Red Riding Hood using
phylogenetic network analysis, Digital Scholarship in the Humanities 31(3):611–636.

**Supervisor: Alexandr Gorbunov**

1. **Solving data analysis / machine learning problems cost-optimally using constraint solvers**

***Abstract:*** Constraint optimization, maching learning, and data mining are today well-established and thriving research fields within computer science. Each of the fields have contributed fundamental techniques and algorithmic solutions that are today routinely applied for addressing hard computational problems in various real-world context. However, the possibilities of harnessing the highly efficient constraint solving technology available today in providing generic and efficient solutions to various machine learning and data mining problems, such as different kinds of classification, structure learning, probabilistic reasoning, and pattern mining tasks, have only recently been realized, and there is plenty of opportunities for developing novel algorithmic solutions with optimality-guarantees to machine learning tasks via employing constraint solving.

References:

1. Montanari, U. *Networks of Constraints: Fundamental Properties and Application to Picture Processing*, Information Science, V. 7, (1974), P. 95–132.
2. Shvetsov I., Telerman V., Ushakov D. *NeMo+: Object-Oriented Constraint Programming Environment Based on Subdefinite Models*, LNCS **1330** (1997).
3. Telerman V., Ushakov D., Sidorov V. *Object-Oriented Constraint Programming Environment NeMo+ and its Applications*, Proc. of the 9th Internat. Conf. on Tools with Artificial Intelligence, ICTAI’97.— IEEE Computer Society, Newport Beach, California, USA, (1997).

**Supervisor: Alexandr Gorbunov**

1. **Human powered hierarchical clustering with relative distance comparisons**

***Abstract:*** The input to a hierarchical clustering algorithm is typically a distance matrix that contains pairwise distances between the data items to be clustered. Commonly this matrix is defined e.g. in terms of the Euclidean distance between feature vectors. However, in some applications computing absolute (Euclidean or other) distances is not feasible. This happens for example if the distance information is collected using a crowd of human annotators. Humans are rather poor at consistently evaluating some notion of absolute semantic distance (on some arbitrary scale) between, say, the contents of two photographs. On the other hand, humans are fairly adept at comparing items relative to each other. The objective of this project is to design and analyse an agglomerative hierarchical clustering algorithm that uses (possibly noisy) relative distance comparisons to compute a clustering dendrogram.

References:

1. BLOCKEEL, H. (2006): Experiment databases: A novel methodology for experimental re-search. Lecture Notes in Computer Science, 3933, 72-85.
2. BLOCKEEL, H. and Vanschoren J. (2007): Experiment Databases: Towards an Improved Ex-perimental Methodology in Machine Learning. Lecture Notes in Computer Science, 4702, to appear.

**Supervisor: Alexandr Gorbunov**

1. **Automating a music composition process**

***Abstract:*** There are numerous methods for automated or algorithmic composition of music. This thesis project will approach this goal from a unique point of view: modelling and implementing the composition process of an actual human composer, who will participate in this project. The work will include conceptualisation and formalisation of the composition process together with the composer, in sufficient detail so that it can be implemented as a computational model, and experimented with. Some background in music is useful, so is interest in computational creativity.

References:

1. KALOUSIS, A. and HILARIO, M. (2000): Building Algorithm Profiles for prior Model Se-lection in Knowledge Discovery Systems. Engineering Intelligent Syst., 8(2).
2. PENG, Y. et al. (2002): Improved Dataset Characterisation for Meta-Learning. Lecture Notes in Computer Science, 2534, 141-152.

**Supervisor: Alexandr Gorbunov**

1. **Investigating machine learning in classic games**

***Abstract:*** One of the early successes in applying machine learning to board games was Tesauro's NeuroGammon. In chess, on the other hand, champion level performance was obtained mainly without machine learning. Recently, reinforcement learning has been used in breakthrought in computer go and poker. This area offers various possible thesis topics, for example focussing on a particular game, or a particular learning technique.

References:

1. VAN SOMEREN, M. (2001): Model Class Selection and Construction: Beyond the Pro-crustean Approach to Machine Learning Applications. Lecture Notes in Computer Sci-ence, 2049, 196-217.
2. WITTEN, I.H. and FRANK, E. (2005): Data Mining: Practical Machine Learning Tools and Techniques (2nd edition). Morgan Kaufmann.

**Supervisor: Alexandr Gorbunov**

1. **Machine Learning Techniques for Social Network Analysis**

***Abstract:*** The present Term Paper explores and describes existing machine learning techniques as well as the possibility of their application to the field of Social Network Analysis.

**Supervisor: Olga A. Tsukanova**

1. **Big Data Technologies for Social Network Analysis**

***Abstract:*** The present Term Paper focuses on various approaches to modeling and analysis of large social networks considering the possible changes in technology stack in an Era of Big Data.

**Supervisor: Olga A. Tsukanova**

1. **Customer behavior management based on geolocation data / Управление поведением потребителей с использованием их данных геолокации.**

***Abstract:*** Since companies have already begun to digitally transform their existing customer loyalty programs into mobile applications, new approaches and possibilities are resulting. By collecting and processing the mobility patterns of every single customer combined with the delivery of appropriate offers and incentives at the right time and the right place, may form the future’s customer approach. The processing of the customers’ location data may reveal new information and insights from the customer buying behavior

References:

* Жуков В. И., Komarov M. M., Приемышев А. А. The Behavior of Consumers in the Shop Based on the Data from "Locastor" System, in: 2018 IEEE 20th Conference on Business Informatics (CBI) Vol. 2: Research-in-Progress Papers and Workshop Papers . IEEE Computer Society, 2018. P. 128-133.
* Komarov M. M., Nemeth A. Customer Behavior Management using Big Data and Location Data, in: XVII Апрельская международная научная конференция по проблемам развития экономики и общества: в 4 кн. / Отв. ред.: Е. Г. Ясин. Кн. 4. М. : Издательский дом НИУ ВШЭ, 2017. P. 527-538.
* Bar-David, R. & Last, M., Context-Aware Location Prediction. In: Big Data Analytics in the Social and Ubiquitous Context: 5th International Workshop on Modeling Social Media, MSM 2014, 5th International Workshop on Mining Ubiquitous and Social Environments, MUSE 2014, and First International Workshop on Machine Learning for Urban Sensor Data, SenseML 2014, Revised Selected Papers. Beersheba: Springer International Publishing, 2016. P. 165-185.
* Brath, R. & Jonker, D., 2015. Graph Analysis and Visualization: Discovering Business Opportunities in Linked Data. Indianapolis: John Wiley & Sons, Inc..
* Chen X., Shi D., Zhao B. and Liu F. Mining Individual Mobility Patterns Based on Location History 2016 IEEE First International Conference on Data Science in Cyberspace (DSC), Changsha, 2016, P. 252-259.
* Costa, T. et al.,. You Are Here: Location Analytics And The Rebirth Of Customer Experience, s.l.: s.n., 2014
* Ester M., Kriegel H.P., Sander J., and Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. in Proceedings of the 2nd International Conference on Knowledge Discovery and Data mining, 1996, vol. 1996, P. 226–231.
* Forrester Research, 2015. Forrester’s North American Consumer Technographics® Retail Survey, s.l.: s.n.
* Fujitsu and EKN white paper. Engaging Connected Consumers through Location Insights. 2015: https://sp.ts.fujitsu.com/dmsp/Publications/public/White%20Paper%20Engaging%20Connected%20Consumers%20through%20Location%20Insights\_Jan%2027.pdf [Accessed 20, July, 2017]
* Nguyen A., Komarov M. M., Moltchanov D. Coverage and network requirements of a “Big Data” flash crowd monitoring system using users’ devices, in: Internet of Things, Smart Spaces, and Next Generation Networks and Systems. 16th International Conference, NEW2AN 2016, and 9th Conference, ruSMART 2016, St. Petersburg, Russia, September 26-28, 2016, Proceedings Т. 9870. Хам : Springer, 2016. P. 372-382.

**Supervisor: Mikhail M. Komarov**

1. **Research on business solutions based on mobility // Исследование бизнеса на основе мобильности**

***Abstract:*** Mobility as a resource should be described with description of particular benefits it can give to a particular business processes or complete solutions. Mobility could be internal for the company (e.g. corporate mobility), could be external as part of the major business processes/or as part of the service utilization for the customers. Mobility with data collection techniques and data storage solutions should also be analyzed and described.

References:

* Жуков В. И., Komarov M. M., Приемышев А. А. The Behavior of Consumers in the Shop Based on the Data from "Locastor" System, in: 2018 IEEE 20th Conference on Business Informatics (CBI) Vol. 2: Research-in-Progress Papers and Workshop Papers . IEEE Computer Society, 2018. P. 128-133
* <https://support.sas.com/resources/papers/proceedings13/053-2013.pdf>
* <https://channels.theinnovationenterprise.com/articles/the-impact-of-mobile-technologies-in-warehousing-distribution>

**Supervisor: Mikhail M. Komarov**

1. **Research on architecture of business solutions based on blockchain // Исследование архитектуры бизнеса на основе блокчейн**

Abstract: Description of blockchain should be provided, analysis of existing solutions should be proposed. Particular business solutions should be analyzed/developed with infrastructure requirements analysis and financial analysis of the solution impact.

References:

* <http://bitcoinbook.cs.princeton.edu/>
* <https://www.nber.org/papers/w22952>

**Supervisor: Mikhail M. Komarov**

1. **Big data in retail // Использование больших данных в ритейле**

Abstract: Research on different aspects of applications of big data for retails should be performed. Products, services which are based on big data should be described.

(Topic will be co-supervised with Leroy Merlin Vostok expert).

References:

* Consumers, Big Data, and Online Tracking in the Retail Industry: A Case Study of Walmart
* Improving Retail Performance with Big Data // Architect’s Guide and Reference Architecture Introduction // Oracle Enterprise Architecture White Paper / February 2015

etc. – reports on retail big data solutions.

1. **Smart parks with big data based solutions // Умные парки с решениями на основе больших данных**

***Abstract:*** Big Data Solutions development for the Smart Park. There should be descriptions of architectures, technologies, customer journey maps proposed. Solution should correlate with the Smart park ideas.

(Topic will be co-supervised with Lancaster University Management School).

References:

* <https://www.connected.community/smart-park>

**Supervisor: Mikhail M. Komarov**

1. **DECISION POINT ANALYSIS IN PROCESS MINING MODELS FOR CLINICAL PATHWAYS IMPROVEMENT**

***Abstract:*** A process model (oriented graph) is a specific and simplified view of the reality that intends to represent what happened as described in the real data. It is like a map that shows the different possible trajectories in the process. A clinical pathway (CP) can be defined as a structured and multidisciplinary care plan used to detail essential steps and timing in the care of patients with a specific clinical problem (Rotter et al., 2010).

Another way to describe a CP is to learn it from patient’s point of view. A CP is the trajectory of different health-related events that happen to a patient in the hospital or another kind of facility.

The purpose here is to determine a set of decision rules, one for each node of the developed process model, that provide the probability of following each outgoing clinical path and which are the most representative of the current event log.

References:

* Thomas Rotter, Leigh Kinsman, Erica L. James, Andreas Machotta, Holger Gothe, Jon Willis, Pamela Snow, and Joachim Kugler. Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs. Cochrane Database of Systematic Reviews, 2010.
* Neesha Jothi, Nur’Aini A. Rashid, and Wahidah Husain. Data mining in healthcare a review. Procedia Computer Science, 72:306 – 313, 2015. ISSN 1877-0509.
* Wil M.P. van der Aalst. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer Publishing Company, Incorporated, 1st edition, 2011.

**Supervisor: Elizaveta Prokofieva**

1. **PROCESS MINING OF BIG EVENT DATA IN HEALTHCARE FACILITIES**

***Abstract:*** Process mining provides the technology to analyze the ever-increasing amounts of event data in modern healthcare facilities: hospitals, polyclinics and other important units.

“Big Event Data” relates process mining to Big Data technologies, providing new capabilities to learn the valuable insights from available resources (event logs). Furthermore, process mining problems can be decomposed in two ways, *case-based decomposition* and *activity-based decomposition*. The research is aimed at applying the process mining techniques to large amounts of data by using the right infrastructure and approach.

References:

* W.M.P. van der Aalst. Extracting Event Data from Databases to Unleash Process Mining. In J.vomBrockeandT.Schmiedel,editors,BPM:DrivingInnovationinaDigitalWorld,pages 105–128. Springer, Berlin, 2015.
* Rojas, E., Munoz-Gama, J., Sepúlveda, M., Capurro, D., Process Mining in Healthcare:A literature review, Journal of Biomedical Informatics (2016).
* Neesha Jothi, Nur’Aini A. Rashid, and Wahidah Husain. Data mining in healthcare a review. Procedia Computer Science, 72:306 – 313, 2015. ISSN 1877-0509.

**Supervisor: Elizaveta Prokofieva**

1. **PROCESS MINING FOR RECOMMENDER STRATEGIES SUPPORT IN HEALTHCARE SETTING**

***Abstract:*** Recommender services exist to provide the most appropriate and relevant content in various settings. A recommendation makes different statements about a set of possible actions in healthcare: doctor’s visits. X-ray procedure, surgery and other ones. Process mining provides the technology to analyze the ever-increasing amounts of event data in modern healthcare facilities: hospitals, polyclinics and other important units. The application of process mining techniques could enable researches to develop more personalized recommender services for patients.

References:

* Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes Authors: Mans, Ronny S., van der Aalst, Wil, Vanwersch, Rob J. B.
* Arias, Michael & Rojas, Eric & Munoz-Gama, Jorge & Sepulveda, Marcos. (2015). A Framework for Recommending Resource Allocation Based on Process Mining. 256.
* W.M.P. van der Aalst. Extracting Event Data from Databases to Unleash Process Mining. In J.vomBrockeandT.Schmiedel,editors,BPM:DrivingInnovationinaDigitalWorld,pages 105–128. Springer, Berlin, 2015.

**Supervisor: Elizaveta Prokofieva**

**50. Ultrametric diffusion as a model of microblogging network evolution /**Ультраметрическая диффузия как модель эволюции микроблоггинговой социальной сети

***Abstract:*** In mathematics, an ultrametric space is a special kind of metric space in which the triangle inequality is replaced with d(x, z) ≤ max{d(x, y), d(y, z)}. Important applications arise in the field of denotational semantics, where points represent a certain amount of information or knowledge. A contraction mapping may then be thought of as a way of approximating the final result of a computation (which can be guaranteed to exist by the Banach fixed point theorem). Similar ideas can be found in domain theory. It is intended to obtain a model for social network evolution using ultrametric diffusion formalism (ultrametric Fokker-Planck equation).

References:

Papadimitriou, Fivos (2013). "Mathematical modelling of land use and landscape complexity with ultrametric topology". Journal of Land Use Science. 8 (2): 234–254.

**Supervisor: Andrey V. Dmitriev**