

Author	Vasily Kornilov
Credits	17
Academic Hours	646
Year of study	1,2
Mode of study	Full-time

1. Course Description

This program establishes the minimum requirements to students' knowledge and skills and determines the content of the discipline and the types of studies and reporting.

The present syllabus is aimed for lecturers, teaching this discipline, teaching assistants and students studying 38.04.05 "Business Informatics" of the master's program "Big Data Systems".

This syllabus meets the standards required by:

- Educational standard of National Research University Higher School;
- Master's educational program "Big Data Systems" for 38.04.05 "Business Informatics";
- HSE curriculum of the educational program for 38.04.05 "Business Informatics", Master's educational program "Big Data Systems", approved in 2016.

The following knowledge and competences are useful for better understanding of the course:

- basic English language, both oral and written,
- basics of mathematical analysis,
- basics of information science,
- basic understanding of data analysis.

This is a compulsory course, it is held during the 1-4 modules of the 1st academic year (10 ECTS credits) and during the 1-3 modules of the 2nd academic year (6 ECTS credits). It includes weekly classroom training in accordance with the approved schedule and unsupervised work of students.

2. Learning Objectives

The objective of the Research seminar is to elaborate skills and experience in development work in the process of students' preparation for theses and graduate qualification papers (master's thesis) of the master's program "Big Data Systems". The course focuses on the research of the areas of technologies and big data appliance, study of the practical work with instrumentation of Big Data, as well as the analysis of the development of technologies.

The main purpose of the Research seminar is to develop academic competences in analysis and evaluation of the impact of new information technologies, including Big Data and related technologies on business performance and its architecture, as well as best practices of the Big Data technologies.

The main objectives of the Research seminar are:

- training students skills in an academic work, including preparation and carrying out scientific projects, writing scientific papers;
- training scientific discussion and presentation of ideas, concepts, research results, projects and research papers;
- training the use of Big Data technologies for scientific activities;
- training methods and skills in scientific forecasting for definition of technological trends in the field of information technologies.

The final goal of the Seminar is to make student's scientific activities being permanent and systematic element of the educational process, to include students into the life of typical scientists, to help learning methodology, technology and tools for research activity.

3. Learning Outcomes

After completing the study of the Research seminar, the student should:

know:

- techniques for analyzing data, such as A/B testing, machine learning and natural language processing;
- big data technologies, like business intelligence, cloud computing and databases;
- visualization tools, such as charts, graphs and other displays of the data.

be able:

- to test, compile, analyze and configure the study methods for their further practical application;
- to master new research methods by self-study;
- to analyze and verify the completeness of information found elsewhere, synthesize and add the information lacking if required
- to organize individual and team research;
- to assess the impact of big data technologies on large enterprises and to suggest effective applications of these technologies
- to form the problem of data analysis based on the description of the business problem;
- to determine requirements for the collection of data to address data analysis tasks;
- to choose a method for solving data analysis problems;
- to use the tools of data analysis and visualization.

have the skills of:

- operational data analysis and data mining;

- forming and solving data mining problems;
- analysis of business information about the company and its environment for decision making;
- analytical reporting using business intelligence tools.

4. Course Plan

The Research seminar is topically divided into 4 parts.

1. Mathematical and technological basis of big data tooling is the research into the peculiarities of appliance technologies to the tasks, connected with the formation of information infrastructure of an organization, new opportunities for analytics and decision-taking.
2. Architectural solutions on the basis of big data for enterprises: transformation of the system of data managing, formation of information assets of an enterprise, systems of data collection about business processes, elaboration of external data, new design of cooperation, system interaction.
3. Big data economics: evaluation of economic efficiency of solutions for enterprise management on the basis of big data technologies, possibilities to use information assets of enterprises, opportunities to use solutions on the basis of unstructured information from various sources in the enterprise administration.
4. Prospects for the development of functionality and spheres of technology appliance.

The general scheme of the Research seminar is shown in table 1. Realization of separate forms does not necessarily coincide with the borders of the relevant modules, nevertheless it is the pointed out forms that are given priority to in every module.

Table 1. Course Plan

1 st year of education	Modules			
	I	II	III	IV
Term paper preparation	Choice of topic and development of term paper general plan	Preparation of scientific and analytical review of term paper subject. Preparation and discussion of term paper project	Term paper execution. Conducting researches connected to appraisal of attained results	Term paper approbation
Types of research seminars	Career-guided lectures given by lecturers of the university and experts in the big data theme, including enterprises-partners	Seminars held with the participation of the university lecturers and big data experts. Trainings on how to compose scientifically analytical reviews	Trainings on term paper execution and project recording. Project evaluation trainings	Seminars held with the participation of the university lecturers and big data experts
2 nd year of education	Modules			
	I	II	III	

Course paper preparation	Choice of subject and development of term paper general plan. Research program discussion and the outline of the Masters Dissertation	Masters research performance. Conducting the research connected with the evaluation of the results obtained	Masters research accomplishment. Masters Dissertation text preparation. Preliminary presentation of the Masters Dissertation
Types of research seminars	Seminars on academic paper preparation held with the participation of the lecturers of the university and experts in the big data theme	Trainings on research organization and performance	Seminars on research work execution. Seminars held with participation of lecturers of the university and experts in the big data theme

The 1st year starts with career-guidance lectures, held by the lecturers and big data experts. They share their personal experience in the research work, introduce students to the procedures of the organization of research projects and some of the results obtained, state tasks, which can provide the basis for the graduation papers topics. In addition, part of the lessons might be given by companies' representatives – the University's partners that process and implement solutions based on the technologies of working with big data.

This form serves to help the students choose the topic of the course paper (with a view to future master's thesis) and form the initial plan of the paper by the end of the first module. The second module is devoted to preparation by the students of course paper project, preparation of the project presentation and discussion of the projects during scientific and research seminar. At this stage, the student shall finally decide on the theme and structure of the course paper. By the end of the third module, the student shall choose the subject and the supervisor and present the course paper project in order to be admitted to further participation in the seminar. During the third module, the students start writing the course paper, which shall include analytical study based on available special academic and analytical literature, electronic sources of information, including available statistical and analytical databases. While analyzing the available materials special attention is given to theoretical approaches and research methods.

Analytical review is due to become the main goal. Trainings on the documentation of the project, research and evaluation of the obtained results are held as well. An essential part of a seminar is dedicated to the development of skills of usage the big data tooling to provide analytical part of work, especially using Internet information. Student start developing data research competences.

During the 4th module students make presentations of the completed works, carry out discussions, correct and complete their course papers on the basis of their discussions. Seminars serve as consultations during this period. The main objective is to help the author to finish the initial text and to lead it to a final look. After that, the papers are peer-reviewed by the big data experts.

The structure of the Seminar of the 2nd year of education is literally the same as that one of the 1st year. It is oriented on the Masters Dissertation preparation.

In the 1st module the 2nd cycle of career-guidance lectures is held by the university lecturers and big data experts. In addition, general structure and demands on how to write the Masters Dissertation are examined. It enables students to choose its topic and the research volume. The choice of the topic and the supervisor has to be made in the course of the first month. At the end of the module, a discussion of the detailed plans of the dissertations is carried out. If a student has failed to present a detailed plan of his dissertation by the end of the module, he will have 2 more weeks to do it during the 1st 2 weeks of the 2nd module.

The 2nd module is dedicated to the discovery of the scope and significance, research program preparation, requirements definition of the results of the research, Masters Dissertation structure formation, preparation of the outline of the Masters dissertation the volume of which cannot exceed 10 pages. The results are discussed in the form of a presentation at the Scientific and Research Seminar.

The main objective for a student in the 3rd module is to work at the dissertation, the core part of which compiles the formation of the hypothesis and the research characteristics, data collecting and analytical research performance. An essential part of the work is the choice, and in some cases it is the creation of a new programming tooling to carry out the data research. At the same time, the lecturers of the university give classes and consultations on the research organization and performance, usage of research and information sources, usage and creation of the research and program tooling based on big data technologies.

In the 4th module students complete the Masters Dissertation preparation and write the primary text for the Masters dissertation. They also get ready for the preliminary dissertation presentation. The preliminary presentation takes place at a seminar being observed by lecturers and experts.

The seminar is supposed to be teamwork between the presenter and the lecturers.

5. Reading List

a. Required

1. Rajkumar Buyya, Rodrigo N. Calheiros and Amir Vahid Dastjerdi. Big Data: Principles and Paradigms. Morgan Kaufmann, 2016. 494 pages. DOI: <https://doi.org/10.1016/C2015-0-04136-3>
2. Steve Williams. Business Intelligence Strategy and Big Data Analytics: A General Management Perspective. Morgan Kaufmann, 2016. 240 pages. DOI: <https://doi.org/10.1016/C2015-0-01169-8>

3. Shen Liu, James McGree, Zongyuan Ge, Yang Xie. Computational and Statistical Methods for Analysing Big Data with Applications. Academic Press, 2016. 206 pages. DOI: <https://doi.org/10.1016/C2015-0-00198-8>
4. Jules J. Berman. Principles and Practice of Big Data: Preparing, Sharing, and Analyzing Complex Information, 2nd Edition. Academic Press, 2018. 480 pages. DOI: <https://doi.org/10.1016/C2017-0-03409-2>

b. Optional

1. IBM Big Data University: <http://bigdatauniversity.com/wpcourses/?cat=19>
2. Big Data Explained | MongoDB: <http://www.mongodb.com/big-data-explained>
3. Learn About Big Data | Intel: <http://www.intel.eu/content/www/eu/en/big-data/learn-about-big-data.html>
4. What is Big Data? | SAS: http://www.sas.com/en_us/insights/big-data/what-is-big-data.html
5. Oracle Technology Network. Big Data: <http://www.oracle.com/technetwork/articles/bigdata/index.html>
6. Big Data Resource Library | TeraData: <http://www.teradata.co.uk/big-data/resources/>

6. Grading System

Students' progress is evaluated at the end of the year. Any failed academic assignment on a research seminar is equal to a usual exam failure.

Students get their credit with due consideration of the results of their research, the level of its novelty, analytical relevancy. Attendance is taken into account as well as active participation in discussions.

After completing the 1st year a student has to grant the following materials:

- a detailed plan of the course paper; analytical review of the bibliography, electronic sources of information, including the data base on the topic of the course work (Module 3), both in electronic and paper form;
- the term paper not exceeding 30 pages in electronic and paper form, and a presentation (7-10 slides) in electronic form to get the final mark (Module 4).

Approximate contents of the course paper:

1. Problem statement. Description of the topical area of the research.
2. Grounding of the research actuality.
3. Scopes, goals and methodological framework of the research.
4. Carrying out of the research.
5. Interpretation of the obtained results.
6. Conclusions.

7. List of bibliography and internet sources (with references in the text).

The materials presented must be the result of individual research work of students, which is carried out by supervision of their Academic Advisers. Students make presentations based on this work and they give them during seminar classes.

Assessment in all forms of current control are set on a 10-point scale.

The final grade (*FG*) for the discipline consists of the following elements:

$$FG = 0.4 * AG + 0.6 * FE$$

Accumulated Grade (*AG*) includes:

- Activity score (*AS*) (preparation and presentation of oral reports, participation in discussions, performing additional tasks)
- Current knowledge control (*CC*) (solving control tasks, tests).

The formula for Accumulated Grade is:

$$AG = 0.6 * AS + 0.4 * CC$$

Final examination (*FE*) includes presentation and defense of the class project and a final test.

The formula for Final Examination Grade is:

$$FE = 0.5 * \text{Grade for the final test} + 0.5 * \text{Grade for the project presentation.}$$

Conversion of the concluding rounded grade to five-point scale grade is done in accordance with the following table:

Table of Grade Accordance

Ten-point Grading Scale	Five-point Grading Scale	
1 - very bad 2 – bad 3 – no pass	Unsatisfactory - 2	FAIL
4 – pass 5 – highly pass	Satisfactory – 3	PASS
6 – good 7 – very good	Good – 4	
8 – almost excellent 9 – excellent 10 – perfect	Excellent – 5	

7. Special Equipment and Software Support

Modern PCs in sufficient quantity, OS MS Windows, MS Office, Internet access, multimedia projector