



Government of Russian Federation
Federal State Autonomous Educational Institution of High
Professional Education
“National Research University Higher School of Economics”

Faculty of Computer Science
School of Data Analysis and Artificial Intelligence

Syllabus for the course
Scientific Seminar "Intelligent Systems and Structural Analysis"

01.04.02 "Applied Mathematics and Informatics" training direction of the "Data
Science" master educational program

Author(s):

Galitsky B.A., bgalitsky@hotmail.com

Ilvovsky D.A., dilvovsky@hse.ru

Makarov I.A., iamakarov@hse.ru

Approved at the meeting of the School of Data Analysis and Artificial Intelligence on

«___»_____ 2018

The head of the Data Analysis and Artificial Intelligence

Kuznetsov S.O. _____

Recommended by the academic council of the educational program "Applied Mathematics
and Informatics" on

«___»_____ 2018, № protocol _____

The academic head of the educational program

Kuznetsov S.O. _____

Moscow, 2018

*This document can not be used by other departments of the university and other
universities without the author's permission.*



1 Scope of use

This program establishes the minimal requirements to students' knowledge and skills and defines the content of the discipline and the types of studies and reports.

The present syllabus is aimed at faculty staff teaching the Scientific Seminar "Intelligent Systems and Structural Analysis", teaching assistants and students studying 01.04.02 "Applied Mathematics and Informatics" of the educational program "Data Science".

This syllabus meets the standards required by:

- Educational standards of Federal State Autonomous Educational Institution of High Professional Education "National Research University "Higher School of Economics";
- Master educational program "Data Science" for 01.04.02 "Applied Mathematics and Informatics";
- University curriculum of the educational program for 01.04.02 "Applied Mathematics and Informatics" training direction of the "Data Science" master program, approved in 2018.

2 Summary

The discipline goal is to develop students' professional skills required for independent analytical work in applied fields of the computer science. Also, this course aims to improve skills of students in developing their research projects related with dialogue systems and chat bots.

This course focuses on analysis of scientific and industrial linguistic system developing and motivates visiting different scientific colloquium at the university, especially at the faculty of computer science.

3 Learning Outcomes

The Scientific Seminar should help students to form the basic skills training to make and present their own research, motivate to engage in the scientific activity.

After completing the study of the Scientific Seminar the student should:

- 1) *know* basic principles of developing task-oriented linguistic dialogue systems;
- 2) *know* fundamental approaches to natural language understanding and dialogue management in the task-oriented dialogue systems;
- 3) *be able to* formulate the task and goals for an independent research and/or scientific programming system development;
- 4) *be able to* prepare a presentation based on his research and/or scientific programming system;

As a result of studying the course student develops the competencies shown at the table below:



Key	Competence	Forms and techniques of the study which lead to developing the competence	Form of control on competence level
SC-2	The ability to create new theories, invent new ways and tools of professional activity.	Seminar classes	<ul style="list-style-type: none">• Attendance and classroom work at the seminar,• preparation of presentation on a scientific article and reporting it at the seminar,• individual project.
SC-3	The ability to independently understand new research methods, changing of the scientific and production profile of their activities.	Seminar classes	
SC-5	The ability to make managerial decisions and is ready to be responsible for them.	Seminar classes	
SC-7	The ability to organize and manage multilateral communication.	Seminar classes	
PC-1	The ability to translate the norms of a healthy lifestyle, to lead by example.	Seminar classes	
PC-2	The ability to translate legal and ethical norms in professional and social activities.	Seminar classes	
PC-5	The ability to consciously choose strategies for interpersonal interaction.	Seminar classes	
PC-7	The ability to build professional activity, business and make choices, guided by the principles of social responsibility.	Seminar classes	
PC-10	The ability to analyze and reproduce the meaning of interdisciplinary texts using the language and tools of applied mathematics.	Seminar classes	
PC-11	The ability to create interdisciplinary texts using the language and apparatus of applied mathematics.	Seminar classes	
PC-12	The ability to publicly represent the results of professional activity (including the use of information technology).	Seminar classes	
PC-13	The ability to carry out targeted multi-criteria search of information on the latest scientific and technological achievements in the Internet and in other sources.	Seminar classes	
PC-14	The ability to describe problems and situations of professional activity, using the language and apparatus of applied mathematics in solving interdisciplinary problems.	Seminar classes	
PC-15	The ability to create, describe and responsibly control the realization of technological requirements and regulatory documents in professional activities.	Seminar classes	



Key	Competence	Forms and techniques of the study which lead to developing the competence	Form of control on competence level
PC -16	The ability to use knowledge in the field of natural sciences, mathematics and computer science, understanding of the basic facts, concepts, principles of theories related to applied mathematics and computer science.	Seminar classes	
PC -17	The ability to build and solve mathematical models in accordance with the direction of training and specialization.	Seminar classes	
PC -18	The ability to understand and apply modern mathematical tools in research and application.	Seminar classes	
PC -19	As a part of the research and production team, the ability to solve the tasks of professional activity in accordance with the profile of training, communicate with experts in other subject areas.	Seminar classes	
PC -20	The ability to apply modern programming languages and data manipulation languages, operating systems, electronic libraries and software packages, network technologies, etc. in research and application activities.	Seminar classes	

4 Place of the Discipline in the Postgraduate Program Structure

This is a compulsory course for specialization "Intelligent Systems and Structural Analysis", 01.04.02 "Applied Mathematics and Informatics".

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

- basic English language, both oral and written;
- basic knowledge on higher mathematics and computer science.

5 Schedule

№	Topic	Total hours	Contact hours	Self-study
			Seminars	
1.	Developing task-oriented chat bots	248	48	200
2.	Pre-defense of the master's thesis	132	20	112
Total		380	68	312



6 Grading System. Guidelines for Knowledge Assessment

Control type	Assessment	1 year				Parameters
		1	2	3	4	
Current	Presentation 1	*				Progress report on the programming project. Speaking time is no more 15 min.
	Programming project		*			Report on the programming project: individual paper report and group presentation.
	Presentation 2			*		Report on the master's thesis. Speaking time is no more 15 min.

The assessments consist only of tasks for the current control.

The current control includes one presentation during the first module based on the current progress in developing programming project which is obligatory. It also includes providing final report on the project and public defense of the project in the form of presentation. Programming project has to be focused on developing dialogue system or chat bot system and can be done in groups up to 4 students. Final reports have to be individual and should contain description of the task, goals, number of working demo-examples and also should explain which part of the project was done by this particular student.

In addition, students should prepare a presentation on the results obtained in the framework of the master's thesis.

The final mark is evaluated like:

$$O_{final} = 0.8 \cdot O_{project} + 0.2 \cdot O_{presentation2}$$

All these marks are evaluated using 10 grade scale. All grades having a fractional part greater than 0.5 are rounded up.

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 Course description

The course consists of lectures on developing modern task-oriented dialogue systems and student presentations on their programming projects. Theoretical part of the course is original and based on the book (to be published in 2018) on modern chat bot architecture. Lectures and preliminary version of the book are available for the students.

1 Developing modern task-oriented chat bots

Student should choose a scientific article dedicated to formal concept analysis (FCA) or areas closely related to FCA such as data mining, information retrieval, knowledge management, data and knowledge engineering, logic, algebra and lattice theory. Then student should analysis chosen article and prepare its presentation for reporting at the Scientific Seminar. Lectures includes following topics:

Lecture 1 *A basic chat bot*

- Building transactional chatbots with Api.ai
- Building FAQ chatbot with Microsoft QnA Maker
- A chatbot with rule-based dialogue management

Lecture 2 *Social Bots*

- Main principles

Lecture 3 *Task-oriented Bots*

- Main principles

Lecture 4 *NL Understanding*

- Introduction to NLP and NLU

Lecture 5 *Assuring chat bot relevance at syntactic level*

- Syntactic Generalization in search and relevance assessment
- Generalizing portions of text
- Generalizing at various levels: From words to paragraphs
- Equivalence transformation on phrases
- Simplified example of generalization of sentences
- From syntax to inductive semantics
- Nearest-neighbor learning of generalizations
- Syntactic generalization-based search engine and its evaluation
- User interface of search engine
- Qualitative evaluation of search
- Evaluation of web search relevance improvement
- Evaluation of product search
- Comparison with other means of search relevance improvement
- Evaluation of text classification problems
- Comparative performance analysis in text classification domains



- Example of recognizing meaningless sentences
- Commercial evaluation of text similarity improvement

Lecture 6 *Q/A for Bots: Semantic headers and semantic skeletons*

Lecture 7 *Learning Discourse-level structures*

- Answering paragraph-size questions
- From sentence-level to paragraph-level generalization
- Rhetoric structures and speech acts as inter-sentence links
- Adapting RST for multi-sentence search
- Adapting Speech Act Theory for multi-sentence search
- Parse thickets and their graph representation
- Equivalence transformation of phrases
- Finding similarity between two paragraphs of text
- How coreferences help search recall
- How rhetoric relation improve search accuracy
- Thicket Phrases and their generalization
- Example of parse thicket
- Generalization of parse thickets
- Generalization for RST arcs
- Generalization for CA arcs
- Computing maximal common sub-PTs
- Architecture of PT processing system
- Evaluation of PT-supported search relevance
- Evaluation settings
- Pair-wise sentence generalization for question-answer similarity
- Single sentence query and answer distributed through multiple sentences
- Query is a paragraph and answer is a paragraph
- Phrase-based and graph-based implementation of generalization
- Comparison of search performance with other studies

Lecture 8 *Building taxonomy and thesaurus for chat bots*

- Improving search relevance by taxonomies
- Must-occur keywords
- Must-occur keywords in a taxonomy
- Constructing relevance score function
- Examples of filtering answers based on taxonomy
- Taxonomy-based algorithm for filtering search results
- Building taxonomies by web mining
- Building taxonomy by generalizing search results
- Practical considerations
- Evaluation of search relevance improvement by taxonomies
- Evaluation settings of search relevance improvement
- Vertical search
- Web search relevance improvement
- Taxonomy-supported search engine in news domain
- Taxonomies for query expansion
- Using search in Similarity component



- Running taxonomy learner

Lecture 9 *Chat bot content processing pipeline*

- From search to personalized recommendations
- A content pipeline and its relevance-related problems Content pipeline architecture
 - Content processing engines
 - Content processing units
 - Harvesting unit
 - Content mining unit Taxonomy unit
 - Opinion mining unit De-duplication unit Search Engine Marketing unit
 - Speech recognition semantics unit
 - Search unit
 - Personalization unit
 - Generalization of texts
 - Simplified example of generalization of sentences
 - Sample generalization between phrases
 - Tree Kernel approach for text similarity
 - Phrase-level generalization
 - Generalization of expressions of interest
 - Personalization algorithm as intersection of likes
 - Mapping categories of interest / taxonomies
 - Defeasible logic programming-based rule engine
 - Content pipeline algorithms
 - Taxonomy construction algorithm
 - De-duplication algorithms Sentiment analysis algorithm
 - Search engine marketing ad construction algorithm

Lecture 10 *Managing Rhetorical Agreement in Dialogue Utterances*

- Communicative Discourse Trees
- Representing rhetorical relations and communicative actions
- Greedy representations for a Q/A pair
- Communicative actions and their generalization
- Generalization for RST relations
- Representing a Request-Response chain
- Classification settings for Request-Response pairs
- Nearest Neighbor graph-based classification
- Thicket Kernel learning for CDT
- Implementation of Rhetorical Agreement classifier
- Discourse Structure-Driven Dialogue Management
- Maintaining cohesive session flow in a chat bot
- Personalized Domain Exploration Scenarios
- Navigation with the Extended Discourse Tree
- Recognizing valid and invalid R-R pairs
- CDT Construction Task
- Managing dialogues and question answering
- Analytical approaches to RR Agreement
- Rhetorical relations and argumentation



Lecture 11 *Discourse-level Dialogue management*

- Finding Answers with Optimal Rhetoric Representation
- Adjusting rhetoric representation of answer to that of a question
- Maintaining a sequence of discourse trees
- Identifying rhetoric correlation
- Building Dialogue Structure from Discourse Tree of a Query
- Maintaining communicative discourse for Q and A
- Learning complement relation

Lecture 12 *Data for chat bot training*

Lecture 13 *Argumentation for chat bot*

8 Educational technologies

The following educational technologies are used in the study process:

- discussion in the classes
- office hours

Methods of Instruction:

Individual and group course projects.

9 Guidelines for Knowledge Assessment

Basic rules for making presentation:

- Uniform style
- Contrast color scheme (optimum – black symbols on a white background)
- Outline
- Absence of points at the end of headers
- Different short slide titles
- Slide numbering
- Highlight keywords
- Font size: 24–54 points (heading), 18–36 items (plain text)
- Avoid overloaded or empty slides
- One idea per slide
- More examples – more clarity
- Conclusion
- Spell-check
- Save a presentation in .PDF format!

Basic rules for writing review:

- The structure includes:
 - * Title – informs us it is a review
 - * Informative Abstract – informs us this is a meta-analysis
 - * Introduction - describes the area and theme of a report



- * Body – describes content, material and methods used in the survey
- * Conclusion – describes results
- * Discussion – includes own opinion on the report
- Do not use informal language. Academic writing should be formal and without slang words.

10 Materials

Resources:

(1) Books on Natural Language Understanding (available online):

- Jurafsky, D., Martin J. H. Speech and Language Processing / – Pearson Prentice Hall, 2009.
- Manning C.D., Schuetze H. Foundations of Statistical Natural Processing / – MIT Press, 1999
- Bird, S., Klein, E., Loper, E. Natural Language Processing with Python / – O'Reilly Media, 2009.
- Chomsky, N. Syntactic structures / Walter de Gruyter, 2002.

(2) Lectures and materials on Wiki seminar resource

http://wiki.cs.hse.ru/%D0%9D%D0%98%D0%A1_%D0%9D%D0%B0%D1%83%D0%BA%D0%B8_%D0%BE_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85_2018/2019

(3) Book about developing chat bot on Wiki seminar resource

http://wiki.cs.hse.ru/%D0%9D%D0%98%D0%A1_%D0%9D%D0%B0%D1%83%D0%BA%D0%B8_%D0%BE_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85_2018/2019

11 Equipment

Projector.