Syllabus

1 Course Description

a. *Title of a Course:* the Scientific Seminar «Intelligent Systems and Structural Analysis».

b. *Pre-requisites:*

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;
- programming basics (real programming experience is eligible);
- algorithms and data structures basics;
- applied graph theory.

c. *Course Type:* compulsory

d. *Abstract:*

The discipline goal is to develop students' professional skills required for independent analytical work in applied fields of the computer science. It consists of two parts.

d.1 Developing task-oriented chat bots

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.

d.2 Methods of analysis and visualization of structured data

Nowadays, structural analysis became a most valuable part of data analysis. Any system has a *structure* which define significant properties of the system. Using mathematical models such as graphs, digraphs, hypergraphs and networks we can analyze structural characteristics of systems. But it is very important for analysts and experts to have an ability to visualize the structure of system. Thus, visualization of networks has grown more and more significant while analysis of networks has applied in more and more domains, especially for big networks like world-wide social, economic and logistics networks, Internet, electronic circuits, neural nets, metabolism networks, etc.

A bachelor of applied mathematics, who works in the field of data analysis should be familiar with as basic approaches and techniques of analysis and visualization of structured data.

2 Learning Objectives

Upon successful completion of this course, students will demonstrate:

1) ability to develop modern task-oriented dialogue systems;
2) ability to design and solve graph-theoretical mathematical models;
3) ability to select and justify appropriate graph drawing method and algorithm
4) ability to use development techniques, skills and tools necessary to network analysis;
5) ability to use development techniques, skills and tools necessary to network visualization;
6) ability to communicate effectively;
7) understanding of professional and ethical responsibility.

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 programing system development;
4) *be able to* prepare a presentation based on his research and/or scientific programing system;
5) *know* a graph-theoretic description of network analysis task and corresponding network visualization requirements;
6) *know* the classification of main network analysis tasks, basic methods and algorithms, most popular software tools;
7) *be able to* use development techniques, skills and tools necessary to network analysis and visualization;
8) *be able to* design and solve graph-theoretical mathematical models;
9) *be able to* select and justify appropriate graph drawing method and algorithm;
10) *possess* reasonably an appropriate project solutions and tools for network analysis workflow.

4 Course Plan:

1 Developing modern task-oriented chat bots

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.

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*

2 Methods of Analysis and Visualization of Structured Data Topic

1 *Solution of LPP* (Link Prediction Problem)

Topic 2 *Reading and discussing articles from "International Symposium on Graph Drawing"* (http://www.graphdrawing.org/)

Each student would choose one article from GD annual almanac, read it and give short talk on the

chosen topic. Expected time of speech: 15 min + 5 min for questions and discussion.

Topic 3 *Large graphs visualization*

Topic 4 *Implementation of SIR model with percolation*

5 Reading List

a. Required

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 _2018/2 019
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
4. Bollobas B. *Modern Graph Theory*. Corrected ed. Springer, 2013. 394 p.
5. *Handbook of Graph Drawing and Visualization*. Edited by Tamassia R. CRC Press, 2013. 862 p.
6. Awesome-network-analysis – A curated list of awesome network analysis resources. Web. (http://github.com/briatte/awesome-network-analysis)
7. The GraphML File Format. Web. (http://graphml.graphdrawing.org)

b. Optional

1. Jackson M.O. *Social and Economic Networks*. Princeton University Press, 2010. 520 p.
2. Yang S., Keller F.B., Zheng L. *Social Network Analysis: Methods and Examples*. SAGE, 2016. 248 p.

3. *Graph Drawing*, 5st – 25th International Symposium, 1997-2007 (See http://dblp.uni-trier.de/db/conf/gd/ and http://graphdrawing.org/symposia.html).
4. The House of Graphs. Web. (http://hog.grinvin.org)
5. Information Resources Management Association (IRMA), *Human-Computer Interaction: Concepts, Methodologies, Tools, and Applications Vol. I-IV*. IGI Global, 2016.

6 Grading System

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.
	Home assignment and in-class assignments			*		

The assessments consist only of tasks for the current control.

6.1 *The current control* of the first part called "Developing task-oriented chat bots" 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: $O_{\text{ppppppp}1} = O_{\text{ppppppppp}}$.

6.2 Students' final grades of the second part called "Methods of Analysis and Visualization of Structured Data" are based on the following activities: *home assignments* at seminars and *in-class assignment.*

Ongoing assessment is delivered as *home assignment* (HA) and *in-class assignments* (ICA). Grades, which are gained by students while ongoing assessment, are parts of the

cumulative grade (CG). Final grade ($OO_{ppppppp2}$) is calculated as follows: $OO_{ppppppp2}$ = CG.

HA1 = 4

points HA2 =

3 points ICA =

3 points

$$OO_{ppppppp2} = HHAA_1 + HHAA_2 + IIIIAA$$

The final mark for the course is evaluated like:

$$Ofinal = 0.6 \cdot Opart1 + 0.4 \cdot Opart2$$

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

7 Guidelines for Knowledge Assessment

<u>Basic rules for making presentation:</u>

- 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!

<u>Basic rules for writing review:</u>

- 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.

8 Methods of Instruction

Individual and group course projects.

9 Special Equipment and Software Support

Projector, computer classes.